The Dissertation Committee for Catherine Soo Rhee Certifies that this is the approved version of the following dissertation:

TRANSCRIPTIONAL AND EPIGENETIC MECHANISMS OF THE FIRST CELL FATE DECISION AND REPROGRAMMING

Committee:

Jonghwan Kim, Supervisor

Haley O. Tucker, Co-Supervisor

Vishwanath R. Iyer

Steven A. Vokes

Edward M. Marcotte

TRANSCRIPTIONAL AND EPIGENETIC MECHANISMS OF THE FIRST CELL FATE DECISION AND REPROGRAMMING

by

Catherine Soo Rhee, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2016

Dedication

I dedicate this work to

my parents, Rhee and Hong, and my sister Chie and her family,

who have raised me to be the person I am today,

for their unconditional love and support;

my fiancé, Vincent, who has been with me since I was in Madison seven years ago,

for being my guardian, with his great support and continuous love.

I couldn't have finished my PhD without them.

Thank you for all of your support along the way.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Jonghwan Kim, for everything he has done for me during my graduate study at UT-Austin. He has given me dreams to look forward to. I thank Dr. Kim for being my mentor in many ways, not only in science, but also in other parts of my life. He has given me more than any other graduate students could get from their supervisors. His guidance and persistent help made this dissertation possible. Without him, I would not be the person I am today. I am so delighted to be his first graduate student. The best decision I made during my PhD training was joining his lab and being his mentee.

I would also like to thank my co-supervisor, Dr. Haley Tucker, for her unwavering support and encouragement. She has given me many valuable comments and suggestions on my projects, along with important lessons about being a good scientist. I thank her for fostering my independence. It has been privilege to work with her. I am pleased to complete this work as her last graduate student.

I would like to thank my committee members, Dr. Vishwanath Iyer, Dr. Steven Vokes, and Dr. Edward Marcotte, for their advice and input throughout the course of my PhD training. I thank them for their precious time to attend my qualified exams and committee meetings.

Thanks to the entire Jonghwan Kim lab and Haley Tucker lab, past and present, for their support, engaging discussions, and providing an enjoyable atmosphere to work in. Enough thanks cannot be given to Dr. Bum-Kyu Lee for his limitless help in completing this work. The influence of Dr. Samuel Beck has also been invaluable.

Finally, I would like to thank my family, Rhee, Hong, Chie, and Baik, in everything from being supportive to always being there for me. The most beautiful thing in this world is to see my parents smiling and knowing that I am the reason behind that smile. I am so delighted to make them smile by achieving this PhD degree. Last, but not least, I would like to thank my soon-to-be husband, Vincent, for who he is. I am thankful for the time I spent with my best friend, Vincent.

TRANSCRIPTIONAL AND EPIGENETIC MECHANISMS OF THE FIRST CELL FATE DECISION AND REPROGRAMMING

Catherine Soo Rhee, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Jonghwan Kim

Co-Supervisor: Haley Tucker

The placenta is a transient but vital organ mediating a myriad of interactions between maternal and embryonic tissues. The cells in the trophectoderm (TE) lineage are responsible for proper implantation, placentation, and immunological functions of the placenta. However, our understanding of molecular mechanisms underlying placentation and TE development is still rudimentary. Deciphering the mechanisms by which key TE-specific transcription factors (TFs) control the first cell fate decision, as well as the maintenance and differentiation of TE, is a prerequisite for understanding early embryonic development and ultimately improving healthy pregnancy.

First, using a combination of functional genomics, bioinformatics, and mouse genetics, I revealed that Arid3a is a critical regulator for controlling the first cell fate decision and placental development. Ectopically expressed Arid3a induces TE-like gene expression programs in embryonic stem (ES) cells. Moreover, Arid3a is not only essential for maintaining self-renewing TS cells, but also for promoting further differentiation of trophoblastic lineages. Consistently, Arid3a-/- mice suffer from severely impaired post-implantation development, resulting in early embryonic lethality. I further showed that Arid3a directly activates TE-specific genes while repressing

pluripotency genes by recruiting HDAC1. Second, I studied the mechanisms underlying TF-mediated conversion of ES to trophoblast stem (TS)-like cells. Upon overexpression of TS cell-specific TFs, Cdx2, Arid3a, and Gata3 (CAG factors) in ES cells, I performed time–course profiling of chromatin accessibility, transcriptomes, and occupancy of these reprogramming factors during reprogramming. Using an integrative analysis, I discovered that CAG factors orchestrate the conversion via a sequential two-step regulation in a timely, ordered manner, with repression of pluripotency genes by decommissioning active enhancers, followed by activation of TS cell-specific genes as pioneer factors that can access closed chromatin.

Taken together, my studies unveiled that Arid3a functions as a pivotal regulator of TE and placental development by regulating the commitment to the first cell fate, as well as by executing TE lineage differentiation. I advanced our understanding of the mechanisms underlying TF-mediated reprogramming of ES to TS-like cells, in particular Arid3a-mediated transcriptional and epigenetic regulation. Thus, my studies will be beneficial for enhancing clinical applications such as disease modeling, drug screening, and regenerative therapies.

Table of Contents

CHAPTER 1: GENERAL INTRODUCTION 1
1.1 Early mammalian embryogenesis 1
Overview of preimplantation development 1
The first cell fate decision: inner cell mass (ICM) and trophectoderm (TE) ..2
Transcriptional regulation 3
Epigenetic regulation 5
Signaling pathways 6
Second lineage specification in preimplantation 7
Importance of placentas 7
1.2 *In vitro* model systems of early embryogenesis 8
Embryonic stem (ES) cells 8
The pluripotency transcriptional network 9
Trophoblast stem (TS) cells 10
Core TS cell-specific transcription factors 12
Cross talk between pluripotency and epigenetic factors in ES and TS cells 12
Extraembryonic endoderm stem (XEN) cells 13
1.3 Reprogramming 14
Transcription (TF)-mediated direct reprogramming 14
Mechanisms of direct reprogramming 15
ES to TS-like cell fate conversion 16
1.4 Aims of the research project 17
Aim 1. To determine the function of Arid3a in early embryogenesis. 18
Aim 2. To examine mechanisms of TF-mediated reprogramming of ES to TS-like cells. 18
CHAPTER 2: ARID3A IS ESSENTIAL FOR EXECUTION OF THE FIRST CELL FATE DECISION VIA DIRECT EMBRYONIC AND EXTREMBRYONIC TRANSCRIPTIONAL REGULATION 20
2.1 INTRODUCTION 20
2.2 MATERIALS AND METHODS 23

2.3 RESULTS 29

Arid3a is dispensable for mouse ES cell self-renewal but critical for normal differentiation 29

Induction of Arid3a in ES cells promotes TE differentiation 31

Induction of Arid3a in ES cells results in stable TS-like cells that are capable of further differentiation to trophoblast lineages 38

Arid3a and Cdx2 act both independently and in concert to antagonize Oct4 and promote ES cells to TS cell *trans*-differentiation 40

Nuclear localization of Arid3a is required for Oct4 repression and TE differentiation.....43

Target occupancy of Arid3a correlates with Arid3a-mediated gene expression 47

Arid3a directly activates TE markers and directly represses pluripotency factors during ES-to-TS cell *trans*-differentiation.....50

Arid3a represses pluripotency-associated genes by recruiting HDACs52

2.4 DISCUSSION 54

CHAPTER 3: ARID3A IS REQUIRED FOR MAMMALIAN PLACENTA DEVELOPMENT 58

3.1 INTRODUCTION 58

3.2 MATERIALS AND METHODS.....59

3.3 RESULTS 62

Arid3a is highly expressed during mouse and human placentation.....62

Loss of ARID3A during early mouse gestation results in intrauterine growth restriction and defects in placental development.....63

Arid3a mutant placentas have structural defects in placental organization 66

Aberrant expression of subtype-specific markers in Arid3a null placentas 66

Activation of the inflammatory response in *Arid3a* null placentas 68

Loss or gain of Arid3a disrupt expression levels of key TS and TGC differentiation markers.....71

Human ARID3A levels are induced by BMP4 and ARID3A OE leads to down-regulation of pluripotency and up-regulation of TE-associated gene expression.....72

3.4 DISCUSSION 73

CHAPTER 4: MECHANISMS OF TRANSCRIPTION FACTOR-MEDIATED DIRECT REPROGRAMMING OF MOUSE EMBRYONIC STEM CELLS TO TROPHOBLAST STEM-LIKE CELLS 78

4.1 INTRODUCTION 78

4.2 MATERIALS AND METHODS....80

4.3 RESULTS 86

Ectopic expression of individual CAG factors in ES cells promotes changes in global gene expression and chromatin landscape toward TS cells....86

CAG factors directly regulate both ES and TS cell-specific genes during ES to TS-like cell reprogramming 89

Each CAG factor binds both open and closed chromatin during the early stage of reprogramming regardless of the presence of canonical motifs 94

Arid3a promotes cell fate conversion by repression of ES cell-specific genes followed by activation of TS cell-specific genes....97

CAG factors deactivates pre-existing ES cell-specific enhancers 101

4.4 DISCUSSION 104

CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS....107

SUMMARY 107

Understanding epigenetic marks and signaling pathways of the ES to TS-like cell reprogramming....108

Generation of human TS cells....109

TS cell transcriptional network....110

APPENDIX....111

Appendix A. RT-qPCR primers....111

Appendix B. ChIP-qPCR Primers....112

Appendix C. Cloning Primers....113

REFERENCES 114

VITA 126

CHAPTER 1: GENERAL INTRODUCTION

1.1 EARLY MAMMALIAN EMBRYOGENESIS

During early embryogenesis, the totipotent single cell zygote undergoes a sequence of cell divisions accompanied by distinct cellular, molecular, and epigenetic changes, ultimately leading to the partitioning of cells in the developing blastocysts according to their cell fates (Hemberger et al., 2009; Tarkowski, 1959; Zernicka-Goetz et al., 2009). These cells develop into three morphologically distinct lineages by the late blastocyst stage of preimplantation (Kelly, 1977; Ralston and Rossant, 2005). The outer layer of the late blastocyst, or trophectoderm (TE), will develop into the placenta and invade the maternal endometrium to form the intervillous space, which links the growing fetus to the mother(Arnold and Robertson, 2009). The inner layer of the blastocyst is the inner cell mass (ICM), which later differentiates into two distinct tissues: the inner epiblast (EPI) cells and an outer layer of primitive endodermal (PE) cells facing the blastocoel (Dyce et al., 1987; Grabarek et al., 2012). While the PE cells are restricted to the yolk sac, EPI cells can give rise to all adult tissues (Arnold and Robertson, 2009). From the aforementioned cell lineages, three types of stem cell models have been established: (1) trophoblast stem (TS) cells from the TE, (2) extraembryonic endoderm stem (XEN) cells from the PE, and (3) embryonic stem (ES) cells from the EPI or earlier stages of the ICM (Cahan and Daley, 2013; Rossant, 2008). These stem cells serve as invaluable *in vitro* model systems that help illuminate the mechanisms underlying early embryogenesis, as well as lineage specifications, and enable stem cell therapies, which may be used to treat numerous disorders and injuries in the future.

OVERVIEW OF PREIMPLANTATION DEVELOPMENT

During the initial rounds of zygotic cleavage, the cells are morphologically identical and are distributed symmetrically within the embryo until compaction at the eight-cell (morula) stage, which is when these cells become adhesive and polarized (Johnson, 2009; Rossant, 2004). Junctional complexes are gradually formed at apicolateral and lateral

sites, followed by polarization of outer cells (Bischoff et al., 2008). Compaction and polarization during the morula stage generate cellular asymmetry, leading to regression of totipotency and formation of the polarized outer and apolar inner compartments—the TE and the ICM, respectively (Dietrich and Hiiragi, 2007). This segregation process is termed the *first cell fate decision*, after which these two groups of cells diverge sharply in terms of transcriptional and epigenetic regulation during development (Johnson and Ziomek, 1981; Rossant and Tam, 2009; Zernicka-Goetz et al., 2009). Mouse and human embryos undergo similar embryonic developmental processes, although the timeline to reach the blastocyst is delayed in humans compared to mouse embryos (Niakan et al., 2012).

Despite intensive studies that have advanced our understanding of embryogenesis over recent years, it is still not well-understood how the first cell fate decision and lineage specification are controlled by signaling pathways, as well as global transcriptional and epigenetic regulatory mechanisms.

THE FIRST CELL FATE DECISION: INNER CELL MASS (ICM) AND TROPHECTODERM (TE)

Both cell polarity and position in the embryo influence the first cell fate decision. Upon blastocyst formation, the cleavage plane and polarization axis are perpendicular, resulting in the formation of inner apolar and outer polar cells (Jedrusik et al., 2008; Suwinska et al., 2008). Inner apolar cells are progenitors of the pluripotent ICM, which can engender all three germ layers, including the mesoderm, endoderm, and ectoderm, whereas the outer polar cells are antecedents of the multipotent TE that can be differentiated into all the cells of the placenta (Takaoka and Hamada, 2012). In addition, cell polarity and cell position cross-regulate one another, as transplantation of inner cells to an outside position results in polarization and adaption to the TE fate. On the other hand, downregulation of key polarity molecules such as aPKC (atypical protein kinase C) and PARD3 (par-3 family cell polarity regulator) promotes allocation of the cells to inside positions (Pauken and Capco, 1999; Plusa et al., 2005). Although segregation of the ICM and the TE becomes apparent as polarization of blastomeres occurs, cells are not

yet fully committed toward ICM or TE lineages at the 16-cell stage (Dietrich and Hiiragi, 2007). Manipulation of the cells at this stage can alter their cell fate; thus, they are still plastic and totipotent. Cell fate is further determined by signaling cascades of environmental cues, followed by changes in transcriptional activities coupled with selective epigenetic marks.

The mechanisms underlying the first cell fate decision are remarkably complex and remain poorly understood. Recent studies identified substantial changes in the transcriptome during the first cell fate decision, suggesting important roles for transcription factor (TF) activities (Moignard and Gottgens, 2014). In addition, microRNAs (miRNAs) and epigenetic regulations also direct lineage specification (Kloosterman and Plasterk, 2006; Yang et al., 2008).

Transcriptional regulation

TFs play crucial roles during the development of the blastocyst. Importantly, some TFs show restricted expression patterns associated with the segregation of the ICM and the TE. For example, Nanog and Oct4 (Pou5f1) are expressed in ICM cells, but not in TE cells, while Cdx2 is exclusively expressed in the TE (Nichols et al., 1998; Strumpf et al., 2005). These mutually exclusively expressed TFs not only serve as markers to distinguish between the ICM and the TE, but also play crucial roles in first cell fate determination.

Oct4 and Cdx2 currently are considered to be the master regulators of the first fate decision based on their respective mutant phenotypes, distinctive gene expression patterns, and differential functions at the molecular level (Niwa et al., 2005). Initially, Oct4 and Nanog were identified as critical factors required for forming and maintaining the ICM because they become restricted to the ICM cells of the blastocyst (Mitsui et al., 2003; Palmieri et al., 1994). However, subsequent genetic studies in mice revealed that both Oct4 and Nanog knockout (KO) embryos form blastocysts without obvious developmental defects (Mitsui et al., 2003; Nichols et al., 1998). Yet, ES cells, the *in vitro* counterpart of the ICM, cannot be easily derived from the ICM of these mutant embryos, nor do they behave as typical ES cells. For example, ES cells from Oct4-KO

embryos tend to differentiate into the extraembryonic trophoblast lineages, whereas ES cells from Nanog-KO embryos are prone to die via apoptosis (Mitsui et al., 2003; Niwa et al., 2000). These results suggested that Oct4 and Nanog are, instead, key regulators of pluripotency and self-renewal (Boyer et al., 2005). Consistent with these observations, recent genome-wide mapping of pluripotency factors, including Oct4, Sox2, and Nanog in ES cells, revealed that these pluripotency ICM factors tightly maintain their expression levels in the interwoven transcriptional gene regulatory network to auto-regulate and control one other through a feed-forward regulatory loop, and that this pluripotency network involves many additional TFs, epigenetic regulators, and miRNAs (Avilion et al., 2003; Boyer et al., 2005).

On the other hand, Cdx2, a key TF in the generation of the TE, was first detected at the eight-cell stage of the embryo (Beck et al., 1995). Cdx2 is exclusively expressed within cells of the outer embryonic segment, while its expression is depleted in the ICM. Cdx2-KO embryos can initiate lineage commitment, but cannot maintain a blastocyst due to a lack of epithelial integrity in their outside cells (Strumpf et al., 2005). This indicates that Cdx2 is crucial for both epithelial identity and trophoblast multipotency. In addition to Cdx2, other TE-specific TFs, including Eomes, Gata3, and Tfap2c, are also solely expressed in the TE (Home et al., 2009; Russ et al., 2000; Weber et al., 2010). Unlike Cdx2, embryos deficient in any of these factors are able to form a blastocyst, but fail after implantation due to defects in more differentiated placental tissues. These results indicate that these factors are required later during TE development. Deletion of another TE-specific gene, Tead4, results in preimplantation lethality, suggesting that its function is upstream of Cdx2 (Nishioka et al., 2008). Tead4-KO embryos fail to activate TE-specific genes, including Cdx2, without an alteration of ICM-specific gene expression, such as Oct4 and Nanog, throughout all blastocysts (Nishioka et al., 2008; Yagi et al., 2007). Thus, Tead4 is crucially required for TE, rather than ICM, development (Nishioka et al., 2008). During the process of TE specification, Tead4 cooperates with its co-activator, Yap1, to activate TE-specific genes (Nishioka et al., 2009). Likewise, Gata3 participates in TE development downstream of Tead4 but in parallel to Cdx2 (Home et al., 2009;

Ralston et al., 2010). In addition to TFs, Fgf signaling, which is mediated by Ras/Erk signaling, is essential for trophoblast expansion and the sequential induction of TE-specific TFs (Maekawa et al., 2005; Yamauchi et al., 1994).

While ICM-specific TFs positively regulate one another, and the same is observed for TE-specific TFs, ICM-specific TFs tend to antagonize TE-specific TFs and vice-versa, which is fundamental to the decision of ICM vs. TE cell fate. For instance, Oct4, Nanog, and Cdx2 directly repress each other to determine whether cells develop into ICM or TE lineages (Strumpf et al., 2005; Zernicka-Goetz et al., 2009). Although our knowledge of the mechanisms underlying ICM and TE segregation has considerably expanded, many questions remain unanswered. For example, how do the key TFs interact with epigenetic regulators, such as histone-modifying enzymes and/or chromatin remodelers, to activate or suppress gene expression? Which additional TFs are essential for blastocyst development, and what are the critical downstream targets of these TFs? How are these factors themselves regulated within the pluripotency or TE-specific network? Answers to these central questions will elucidate the fundamental mechanisms underlying pluripotency of the ICM and multipotency of the TE lineages, as well as advance stem cell-based future cell therapies. My thesis project has focused on tackling several of these fundamental questions.

Epigenetic regulation

From zygote to blastocyst formation, the epigenome, including DNA methylation and histone modifications, changes dramatically, and these changes influence developmental potential (Morgan et al., 2005; Rougier et al., 1998). Asymmetric epigenetic marks are tightly associated with segregation between the ICM and TE (Reik et al., 2003). Recently, DNA methylation and histone modifications have been shown to be indispensable for growth and differentiation of the extraembryonic lineages, suggesting that such modifications, are essential for proper embryonic development (Santos and Dean, 2004; Zheng et al., 2016). In particular, histone methylation signatures such as H3K4me3 and H3K27me3, are enriched at the promoters of ICM-specific genes

exclusively expressed in ICM- and TE-specific genes that are highly expressed in TE. (Li, 2002) Distribution of H3K27me3 and H3K9me3 in ICM and TE are correlated with lineage specification to embryonic vs. extraembryonic tissues (Rugg-Gunn et al., 2010). For example, a histone methyltransferase enzyme, ESET, maintains pluripotency of the ICM through catalyzing a repressive H3K9me3 mark at the promoter of Cdx2 (Carlson et al., 1992; Yeap et al., 2009; Yuan et al., 2009). In contrast, in the TE lineage, methyltransferase Suv39h mediates H3K9me3 at regulatory regions of ICM-specific genes (Alder et al., 2010; Rugg-Gunn et al., 2010). Each of these histone modifications cooperates with developmental-specific TFs to mediate spatial and temporal expression of lineage-specific genes.

Signaling pathways

The Hippo signaling pathway regulates cell fate during embryonic development in position-, cell-polarity-, and density-dependent manners (Harvey and Hariharan, 2012). After the 16-cell stage, Tead4, a downstream effector of the Hippo signaling, is active in outer polar cells to promote TE development (Nishioka et al., 2008). In contrast, cell–cell interactions in the inner apolar cells activate the Hippo pathway, which sequesters Yap1 in the cytoplasm, prevents co-activation with Tead4, and promotes the ICM fate. Although Hippo signaling is spatially regulated at the 16-cell stage, segregation of ICM and TE fate is not clearly patterned until the 32-cell stage. This may be due to another signaling pathway, the Notch pathway, which activates TE-specific genes. A key component of the Notch signaling, RBPJ (also known as CBF1), has stochastic activation up to the morula stage and gradually becomes restricted to the outer cells around the 32-cell stage (Cormier et al., 2004). Thus, the Notch and Hippo pathways cooperate to establish segregation of the ICM and TE, which is clearly observed after the 32-cell stage (Rayon et al., 2014).

Ultimately, crosstalk between multiple signaling pathways – Hippo, Notch, as well as the MAPK pathway – helps regulate cell fate determination in preimplantation embryos. However, further studies are needed to determine other regulatory inputs to lineage

specification at this critical developmental time point to fully understand the mechanisms underlying the first cell fate decision.

SECOND LINEAGE SPECIFICATION IN PREIMPLANTATION

Shortly after the first cell fate decision, ICM cells differentiate into EPI and PE lineages, which is known as the second lineage specification, and this is mediated through FGF/MAPK signaling and differential expression of key TFs, such as Nanog and Gata6 (Baussano et al., 2006; Koutsourakis et al., 1999; Morrisey et al., 1998). FGF signaling orchestrates reciprocal repression between TFs specific for the EPI and PE lineages (Frum and Ralston, 2015). Specifically, Nanog and Gata6, present in all ICM cells until the 32-cell stage, become restricted to either the EPI or PE, respectively. Other TFs, including Gata4 and Sox17, also contribute to the specification of EPI vs. PE lineage within the mouse ICM (Chazaud et al., 2006; Niakan et al., 2010). The expression profiles of these TFs are conserved between human and mouse (Artus and Chazaud, 2014). However, unlike PE formation in the mouse, human embryos are not dependent upon FGF signaling, implying alternative undiscovered signaling pathways might be likely critical for second fate specification in human embryos.

IMPORTANCE OF PLACENTAS

The TE ultimately differentiates into all the specialized cells of the placenta, which nourishes and safeguards the fetus during pregnancy by mediating exchange of nutrients, gases, and waste between mother and fetus, immune protection of the fetus, and secretion of growth factors and pregnancy-related hormones (Cross et al., 1994; Rossant and Cross, 2001). Placental health is essential for a successful pregnancy. In mice, placental tissues are composed of multiple specialized cell types, including trophoblast giant cells, spongiotrophoblasts, and syncytial trophoblasts (Cross, 2005; Rossant and Cross, 2001). The failure of normal TE development can cause mortality and developmental defects not only during pregnancy, but also after birth (Farquharson et al., 2005). The molecular basis of preimplantation is poorly understood despite its crucial importance in pregnancy

health. Thus, the discovery of biomarkers of early placental development will not only serve as diagnostic markers in gynecology, but also will catalyze the development of treatments to save lives of mothers and their fetuses from pregnancy-related disorders.

1.2 *IN VITRO* MODEL SYSTEMS OF EARLY EMBRYOGENESIS

Stem cells are self-renewing (meaning they can renew themselves indefinitely) and either multipotent or pluripotent (meaning they can differentiate into other, more specialized cell types). All blastocyst lineages discussed earlier can be used to derive stem cells for tissue culture; these cells can indefinitely self-renew under appropriate culture conditions and be induced to differentiate *in vitro* or incorporate into the developing embryos *ex vivo*.

EMBRYONIC STEM (ES) CELLS

Embryonic stem (ES) cells were first derived from the ICM of mouse blastocysts (Evans and Kaufman, 1981; Thomson et al., 1998). To maintain their self-renewal and pluripotency, they are cultured either on feeder layers of mouse embryonic fibroblasts (MEFs) or without feeders in the presence of leukemia inhibitory factor (LIF) or under serum-free conditions supplemented with Mek and GSK3 kinase inhibitors (this latter condition is termed "2i") (Evans and Kaufman, 1981; Silva et al., 2008). The combination of 2i and LIF sustains the self-renewing "ground state"; ie, a basal proliferative state that is free of epigenetic restriction and has minimal requirements for extrinsic stimuli. Recent intensive studies have revealed that mouse ES cells express many critical factors including stage-specific embryonic antigens (SSEA-1 and SSEA-3) and TFs that are required for early ICM development such as Oct4, Sox2, and Nanog (Chambers et al., 2003; Loh et al., 2006). ES cells are karyotypically stable, capable of generating chimeras and transgenic mice, and able to form the ectoderm, endoderm, and mesoderm in culture, but they cannot contribute to the TE lineages of the placenta (Yu and Thomson, 2008).

Several years after mouse ES cells were established, human ES cell lines have also been derived, which share some conserved features (Thomson et al., 1998). Similar to mouse ES cells, human ES cells require an OCT4/SOX2/NANOG regulatory module (Boyer et al., 2005). They also express surface markers such as SSEA-3, SSEA-4, and TRA-1, and other factors including epigenetic regulators, miRNAs, and signaling molecules that played important roles in mouse ES cells. However, key differences exist. For example, human ES cells do not require the LIF growth factor for self-renewal (Hirai et al., 2011; Niwa et al., 1998). More importantly, human ES cells possessed transcriptional and epigenetic profiles more similar to those of mouse epiblast stem cells (EpiSC, derived from epiblast of inner cell mass) compared to mouse ES cells (Hanna et al., 2010). Interestingly, some studies, although controversial, suggested that human ES cells are able to differentiate into TE/TS-like cells under appropriate culture conditions (Bernardo et al., 2011; Li et al., 2013; Xu et al., 2002). Thus, further studies establishing interspecies differences are required and observations from mouse ES cells cannot be uniformly extrapolated to human ES cells.

The pluripotency transcriptional network

Genetic studies in combination with high-throughput knockdown (KD) screening revealed that many critical regulators (including sequence-specific TFs, chromatin modifying enzymes, chromatin remodelers, miRNAs, and non-coding RNAs) are indispensable for maintaining self-renewal and pluripotency (Ding et al., 2009; Zhang et al., 2006). Genome-wide target mapping of sequence-specific TFs has unveiled the core transcriptional network of pluripotency factors, composed of Oct4, Sox2, and Nanog, which are essential, as well as other TFs with partially redundant activities such as Klf4, Esrrb, and Tbx3 (Chen et al., 2008; Kim et al., 2008; Loh et al., 2006; Orkin et al., 2008).

Oct4 is a key master regulator that sits atop of the hierarchy of the pluripotency regulatory network. Self-renewal and developmental potential of ES cells are both highly sensitive to either up- or down-regulation of Oct4 expression. For example, Oct4-High cells are more prone to differentiate into the primitive endoderm, whereas Oct4-Low cells

tend to differentiate into TE lineages (Niwa et al., 2000). Sox2 is a well-known interaction partner of Oct4, and alteration of Sox2 levels also disrupts the pluripotency of ES cells. For instance, both Oct4- and Sox2-KO ES cells cannot self-renew and tend to trans-differentiate/reprogram to TE lineages (Hay et al., 2004). In the pluripotency network, Oct4 and Sox2 positively regulate each other and co-occupy the vast majority of active enhancers of ES cell-specific genes to sustain the self-renewing status of ES cells.

In addition to the activation of pluripotency genes, development-related genes must be suppressed to maintain self-renewal in ES cells. Some pluripotency factors may repress development-related genes by occupying regulatory regions. Importantly, multiple factors, including histone modifying enzymes and histone remodeling complexes such as Hdac1, Mta1, and Mbd3 (the NuRD complex) as well as the Polycomb complex (PRC1 and PRC2) and SWI/SNF complex (Brg1 and Baf155), are also implicated in developmental gene repression (Chen and Dent, 2014; Lessard and Crabtree, 2010). In particular, most developmental genes are marked by a bivalent histone signature consisting of both H3K4me3 and H3K27me3, which is considered to be the "poised status" for rapid activation upon differentiation (Bernstein et al., 2006). In summary, key ES cell-specific TFs orchestrate transcriptional gene activation and repression in conjunction with chromatin regulators within the pluripotency network to sustain self-renewal and pluripotency of ES cells.

TROPHOBLAST STEM (TS) CELLS

TS cells can be isolated from the polar TE of mouse embryos and maintained in a proliferative and undifferentiated state *in vitro* in the presence of Fgf4 (fibroblast growth factor 4) with its cofactor, heparin, and a feeder layer of MEFs (Tanaka et al., 1998). TS cells are self-renewing and multipotent. Upon withdrawal of Fgf4, the vast majority of cells become terminally differentiated into trophoblast giant cells, while a small portion of cells differentiate into other trophoblast lineages, including spongiotrophoblasts and labyrinthine cells. Following their injection into an early embryo, TS cells can participate in the development of chimeras in which they contribute exclusively to trophoblastic

components of the placenta and parietal yolk sac, in contrast to ES cells, whose contributions are restricted to the three germ layers (Tam and Rossant, 2003).

Multiple TFs, serum factors, and signaling pathways, such as Ras/Mapk signaling, maintain the self-renewal and multipotency of TS cells by upregulating TS cell-specific TFs such as Eomes (which only requires one of the two signaling pathways) and Cdx2 (which requires both) (Lu et al., 2008). In drastic contrast to the roles of ICM and ES cells, placental biology, including TE differentiation, is poorly understood despite the critical importance of trophoblast lineage development for successful mammalian reproduction. Therefore, it is vital to elucidate critical factors and mechanisms underlying TE differentiation and placental biology. Importantly, human TS cells are not yet available, mainly due to ethical issues, such as the necessity of destroying human embryos.

Interestingly, spontaneous differentiation of human ES cells to the TE lineage, which is called trans-differentiation, can occur infrequently. The most common method to induce trans-differentiation is by BMP4 treatment of ES cells (Xu et al., 2002). BMP4 down-regulates the expression of OCT4, SOX2, and NANOG in ES cells, thus promoting their conversion to TS-like cells through the induction of TS cell-specific TFs, including TEAD4 and GATA3. Despite dramatic changes in cell morphology and global gene expression that are similar to the human TE, these TS-like cells are mixed cultures that express TE and mesodermal and vascular endothelial cell markers, suggesting that BMP4 treatment alone is insufficient to induce complete trans-differentiation toward TE lineage (Bernardo et al., 2011). Although as yet not fully successful, diverse attempts have been made to induce human TS-like cells from ES cells by inhibiting the Activin/Nodal pathway or by the addition of Fgf2 (Sarkar et al., 2015; Yu et al., 2011). This is unfortunate, as it is essential to develop a well-characterized model of human TS cell differentiation to understand early human placentation and inform therapeutic interventions for pathological pregnancies.

Core TS cell-specific transcription factors

A group of key TS cell-specific TFs, including Tead4, Cdx2, Gata3, Eomes, Elf5, and Tfap2c, positively regulate one another to orchestrate TS cell-specific gene expression (Kuckenberg et al., 2010; Ralston et al., 2010). In contrast to ES cells, TS cells do not require the LIF-Stat3 signaling pathway; instead, they need Ras/Mapk and Nodal signaling to maintain self-renewal (Roberts and Fisher, 2011). Notably, depletion of any one of a subset of TFs (e.g., Tead4, Gata3, Elf5, Tfap2c, and Ets2) leads to lethality of the mouse embryo due to defects in proper TE differentiation (Auman et al., 2002; Donnison et al., 2005; Home et al., 2009; Nishioka et al., 2008; Okada et al., 2007). This further suggests key roles for these TFs in placental biology. Ectopic expression of Gata3 and Elf5 in TS cells abrogates self-renewal of TS cells, indicating that the precise levels of some of these TS cell-specific TFs are crucial for the balance between self-renewal and differentiation.

Genetic studies in mice further revealed a hierarchical regulation among the TS cell-specific TFs. Tead4, the most upstream factor, controls the levels of Cdx2, Eomes, and Elf5 (Latos et al., 2015; Niwa et al., 2005; Ralston et al., 2010). Functional and mechanistic studies of Cdx2 further disclosed that Cdx2 not only auto-regulates itself but also is controlled by multiple TS cell-specific TFs, such as Gata3. Unfortunately, only a few TS cell-specific regulators have been identified, exhaustive mapping of genome-wide DNA binding sites of TS cell-specific TFs has not been achieved, and epigenetic regulation of TE gene expression remains relatively uncharacterized. Many questions remain unanswered, including: How do these TS cell-specific TFs control self-renewal and modulate differentiation of TS cells toward more specialized placental cell types? How do they interact with their DNA targets, histone modifying enzymes, and remodelers? These central questions constitute a major challenge for future investigation.

CROSS TALK BETWEEN PLURIPOTENCY AND EPIGENETIC FACTORS IN ES AND TS CELLS

The balance of pluripotency TFs and epigenetic modifications is critical to maintain stemness in ES cells. The complexes of epigenetic regulators, such as Polycomb (Eed,

Ezh2, and Suz12) and MLL (Wdr5 and Ash2l), are positively regulated by Oct4/Sox2/Nanog (Ang et al., 2011; Chamberlain et al., 2008; Pasini et al., 2007). Interestingly, both the Polycomb and MLL complexes are required to keep lineage-specific genes silenced. Genome-wide DNA methylation patterns mediated by Tet proteins and DNA methyltransferases (DNMTs) also maintain the ES cell identity (Cimmino et al., 2011). Thus, crosstalk between the core pluripotency factors and numerous epigenetic regulators is required for maintaining ES cell self-renewal, as well as for ensuring proper cell differentiation.

Unlike ES cells, only a few histone modifications have been globally investigated in TS cells. Among those, the levels of H3K27me3 appear to be globally lower in TS cells than in ES cells (Chuong et al., 2013). This is probably linked to the low expression levels of Eed, a component of the Polycomb complex, responsible for H3K27me3 signatures. In addition, ChIP-seq analyses have revealed that few genomic regions in TS cells that are enriched for H3K27me3 are only rarely localized near transcription start sites. Moreover, H3K27me3/H3K4me3 bivalent domains, which restrain the expression of developmental genes in ES cells, are rather rare in TS cells (Chuong et al., 2013). However, a more recent study showed the existence of bivalency in TS cells (Liu et al., 2016). Thus, further investigation is needed to understand TS cell biology on an epigenetic level.

EXTRAEMBRYONIC ENDODERM STEM (XEN) CELLS

For many years, mouse ES and TS cells have been used as model systems for ICM/EPI or TE biology in order to study mechanisms underlying embryonic and placental development, respectively (Evans and Kaufman, 1981; Tanaka et al., 1998). Recently, XEN cells have been successfully derived from mouse blastocysts at E3.5 (Niakan et al., 2013). As with other stem cells, they can propagate indefinitely *in vitro*, while maintaining their ability to contribute to extraembryonic endodermal lineages after injection into blastocysts *ex vivo*. Although Gata6 and Gata4 are known to be the key XEN-specific TFs, these TFs are not homogeneously expressed within populations of

XEN cells (Capo-Chichi et al., 2005; Fujikura et al., 2002). It still remains unclear whether this heterogeneity represents distinct XEN cell types. Notably, isolated XEN cells are emerging as a useful stem cell model to understand the convergence of their signaling and transcriptional control of cell fate specification and differentiation, and thus provide another layer of control in developmental biology.

1.3 REPROGRAMMING

Cellular reprogramming is the process that allows conversion of one specific cell type to another. Nuclear reprogramming, first established using frog embryos, ultimately enabled animal cloning using somatic cell nuclear transfer (Gurdon and Melton, 2008; Wilmut et al., 2002). This technique allows de-differentiation of terminally differentiated cells into a stem-like state, refuting the principle of irreversible cellular differentiation. Though powerful, nuclear reprogramming has extremely low efficiency (Blelloch et al., 2006). On the other hand, direct reprogramming of one cell type by delivery of critical molecules, such as TFs, miRNAs, and chemicals, into target cell types can cause dramatic changes in morphology, global gene expression, and epigenetic marks (Hanna et al., 2008; Wapinski et al., 2013). This technique is more efficient than nuclear reprogramming, though its mechanisms remain incompletely characterized.

TRANSCRIPTION (TF)-MEDIATED DIRECT REPROGRAMMING

The first instance of successful direct reprogramming was the conversion of fibroblasts into muscle cells by ectopically expressing the MyoD TF (Davis et al., 1987). More recently, Yamanaka's group in Japan established a protocol for the generation of "induced" pluripotent stem (iPS) cells by the overexpression of a handful of TFs (Oct4, Sox2, Klf4, and Myc) in fibroblasts (Takahashi and Yamanaka, 2006). Like ICM-derived ES cells, iPS cells are self-renewing and pluripotent. Their success overcomes ethical issues in deriving ES cells from developing human embryos, thereby paving the way to develop future patient-specific stem cell therapy. For instance, patient-specific iPS cells may be generated from fibroblasts of a patient donor (e.g., a sufferer of Alzheimer's

disease) followed by the correction of genomic mutations by genome editing approaches. These genome-edited cells can be re-introduced to patients for patient-specific stem cell therapy. In addition to iPS cells, numerous studies have recently reported defined TF-mediated cell type conversions in a broad range of cell types, including fully functional induced neural (iN) cells and induced cardiomyocytes (iCMs) (Ieda et al., 2010; Wapinski et al., 2013).

Mouse ES cells can be directly reprogrammed into extraembryonic lineages by the overexpression of single TFs, such as Cdx2 or Gata6, to derive TS- and XEN-like cells, respectively (Niwa et al., 2005; Shimosato et al., 2007). The morphology and global gene expression patterns of TS- and XEN-like cells are highly similar to their counterpart genuine cells, suggesting that TF-mediated reprogrammed cells may acquire proper characteristics *in vitro*. Although TF-mediated reprogramming from numerous cell types is possible, substantial obstacles remain, such as the inconsistencies in differentiation potential. This may suggest that there are still undiscovered TFs and/or epigenetic regulators necessary to augment the efficiency and speed of reprogramming. In addition, many protocols available now have been developed from the mouse model, and it is essential that they be translated to the human system. Notably, it has become possible to reprogram mouse fibroblasts into induced trophoblast stem (iTS) cells by introducing a handful of TFs, including Eomes, Gata3, Tfap2c, and Ets2 or Myc (Benchetrit et al., 2015; Kubaczka et al., 2015). This new methodology will allow us to generate human iTS cells in the near future.

MECHANISMS OF DIRECT REPROGRAMMING

Reprogramming requires that cell-type specific genes of the original cell type must be repressed, while those of the target cell type must be activated. The mechanism of how this occurs is poorly characterized, particularly the repression aspect. Reprogramming to iPS cells can be broadly divided into two gene activation phases: a long stochastic phase followed by a shorter deterministic phase (Stadtfeld et al., 2008). At first, cells undergo increased proliferation and alteration in histone modification to initiate mesenchymal-to-

epithelial transition. Later, cells enter an intermediate phase followed by long latency of the process. In this late phase, cells stochastically undergo an epigenetic reset, as well as activation of their target-specific transcriptional network, which stabilizes the reprogrammed cells. Another well-established gene activation mechanism during reprogramming is the ability by which ectopically expressed TFs act as "pioneer" factors (Zaret and Carroll, 2011). Pioneer factors initially bind to closed chromatin of genes specific to the target cell type. Once bound, pioneer factors subsequently interact with chromatin modifying enzymes or remodeling complexes to convert closed into open chromatin, thus de-repressing target cell-specific genes. For example, early in reprogramming, Oct4, Sox2, and Klf4 function as pioneer factors by accessing closed chromatin of distal regulatory elements of pluripotency genes, such as Esrrb and Sall4 (Soufi et al., 2015). Ascl1, a TF capable of converting fibroblasts to iN cells, also works as a pioneer factor, showing that this concept extends beyond the reprogramming of fibroblasts to iPS cells (Wapinski et al., 2013).

Interesting insights have been gained into the process of reprogramming, including the findings that cells undergo defined sequential molecular events in an apparently stochastic manner, and that these events are influenced by the choice and number of TFs, as well as the starting cell type. However, many questions remain unanswered; in particular, how do ectopically expressed TFs mediate gene repression? This missing mechanism is part of what my research has focused on.

ES TO TS-LIKE CELL FATE CONVERSION

ES cells can be trans-differentiated (a process also known as reprogramming) to TS-like cells that are highly similar to genuine multipotent TS cells with respect to morphology and global gene expression (Kuckenberg et al., 2011). This represents an informative *in vitro* model for the investigation of factors and the mechanisms underlying the first cell fate decision. Several methods promoting ES to TS-like cell conversion have been reported, including KD of Oct4 and induction of trophoblast-specific TFs in ES cells (Niwa et al., 2000). Finally, ectopic expression of components implicated in the

Ras/Erk signaling pathway also induces conversion of ES to TS-like cells (Lu et al., 2008). However, the mechanisms by which ES cell-specific genes are repressed while TS cell-specific genes are activated by these diverse manipulations remain elusive.

Unfortunately, TS-like cells obtained through KD of Oct4 or induction of TE-specific TFs in ES cells thus far do not possess all properties of genuine TS cells. However, these TS-like cells do experience upregulation of core TS cell-specific regulators, and they can to be incorporated into the TE of developing embryos, where they contribute to placental lineages *ex vivo* (Niwa et al., 2005; Rhee et al., 2014). Consistent with this, recent work performed by our own group has shown that the OE of a single factor, Arid3a, a novel player in trophoblast lineage development, causes trans-differentiation of ES cells to TS-like cells, even under ES cell culture conditions (Rhee et al., 2014). Arid3a achieves this by suppressing ES cell core pluripotency factors and activating TS cell-specific genes. Arid3a OE cells also were successfully incorporated into the TE of developing embryos, indicating that they perform equivalently to bona fide TS cells at least in a mouse model system. Thus, ES to TS-like cell reprogramming can serve as a good model for uncovering mechanisms underlying the first cell fate decision, direct reprogramming, and early trophoblast lineage specification.

1.4 AIMS OF THE RESEARCH PROJECT

The placenta is a transient but vital organ mediating a myriad of interactions between maternal and embryonic tissues. The cells in the TE lineage are responsible for proper implantation, placentation, and the vascular and immunological functions of the placenta. However, our understanding of molecular mechanisms essential for placentation and TE development remain unclear. We contend that elucidation of mechanisms utilized by key TE -specific TFs directing the first cell fate decision as well as maintenance and differentiation of TE is prerequisite to improve understanding of early embryonic development and will ultimately have a positive impact on pregnancy health. In addition, elucidating the mechanisms of TF-mediated trans-differentiation/direct reprogramming of ES to TS-like cells can enhance our understanding of basic transcriptional and epigenetic

regulation that govern cell fate choice not only during reprogramming but also normal placental development. Studies of cell fate change, such as from ES to TS-like cell state will enhance clinical applications such as disease modeling, drug screening, and regenerative therapies. Using a combination of functional genomics, bioinformatics, and mouse genetics, my thesis project has undertaken the following Specific Aims:

AIM 1. TO DETERMINE THE FUNCTION OF ARID3A IN EARLY EMBRYOGENESIS.

Arid3a had recently been identified as a member of the ES cell pluripotency network (Wang et al., 2006). We found that Arid3a is moderately expressed in resting ES cells, whereas its expression is significantly increased upon differentiation. In particular, Arid3a is highly expressed in extraembryonic tissues which eventually give rise to the placenta, suggesting a putative role in placental development. Consistent with this, prior loss-of-function studies of Arid3a in mice revealed a lethal phenotype of null embryos prior to E13.5, indicating that it may have an important function during early embryonic development (Webb et al., 2011). Thus, we sought to determine if Arid3a contributed to the first cell fate decision. We examined the action of Arid3a in TE specification by testing whether Arid3a OE-ES cells can establish the TE lineage *in vitro* and *ex vivo*. We also examined the mechanisms by which Arid3a regulates the first cell fate decision by investigating its localization and downstream targets in *in vitro* model systems. Lastly, we determined the functions of Arid3a in placental development in mice, by employing loss-of-function studies in conventional KO mice, and in human, via analysis of BMP4-induced ES to TS-like conversion. These studies constitute Chapters 2 and 3 of the thesis.

AIM 2. TO EXAMINE MECHANISMS OF TF-MEDIATED REPROGRAMMING OF ES TO TS-LIKE CELLS.

TS cell-specific reprogramming factors for ES to TS-like cell conversion are well established as being instrumental in trophoblast differentiation and placental development *in vivo*. However, their mechanisms of gene regulation during the reprogramming are heretofore uncharacterized. Thus, we used three TS cell-specific reprogramming factors,

Cdx2, Arid3a, and Gata3, consequently designated as “CAG factors” to elucidate the mechanisms of TF-mediated reprogramming of ES towards TS-like cells. By examining global mRNA expression changes via RNA-seq, the dynamics of target occupancy utilization via ChIP-seq, and changes in chromatin landscapes via ATAC-seq, we elucidated a novel “two-step, repression and activation” mechanisms for reprogramming of ES to TS-like cells using CAG factors. These studies are presented in Chapter 4 of this dissertation.

CHAPTER 2: ARID3A IS ESSENTIAL FOR EXECUTION OF THE FIRST CELL FATE DECISION VIA DIRECT EMBRYONIC AND EXTREMBRYONIC TRANSCRIPTIONAL REGULATION

2.1 INTRODUCTION

Transformation of a totipotent zygote into a multilineage blastocyst encompasses a series of cellular, morphological, and molecular events required for implantation and development into an offspring (Zernicka-Goetz et al., 2009). Three distinct cellular lineages—the trophectoderm (TE), the epiblast, and the primitive endoderm—comprise the early developing embryo. Two sequential cell fate decisions during blastocyst formation are required to establish these lineages. The first cell fate decision, occurring at the 16- to 32-cell stage, leads to segregation of the TE and inner cell mass (ICM). The TE is required for implantation into the uterus and formation of the placenta, whereas the cells in the ICM are pluripotent and have the capacity to give rise to all tissues and organs in the body (Niwa, 2007). The second cell fate decision controls division of the ICM into the epiblast and the primitive endoderm.

Since embryonic stem (ES) cells are derived from the ICM (Evans and Kaufman, 1981), intensive studies have identified crucial ES cell ‘‘core’’ transcription factors (TFs) such as Oct4, Sox2, and Nanog that are central to the construction of a sophisticated transcriptional regulatory circuit termed the ‘‘core pluripotency network.’’ TFs within the circuitry work in concert to allow ES cells to self-renew while maintaining pluripotency, primarily by activating pluripotency-associated genes and repressing lineage-specific TFs(Kim et al., 2008). In contrast, only a few TFs required for TE specification have been identified, and much less is understood as to how they orchestrate the segregation of the ICM and TE during the first cell fate decision(Zernicka-Goetz et al., 2009). The first

---

Reproduced in part from Catherine Rhee, Bum-Kyu Lee, Samuel Beck, Azeen Anjum, Kendra R. Cook, Melissa Popowski, Haley O. Tucker, and Jonghwan Kim(2014). Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes Dev.* 28(20):2219-2232. (Contributions: C.R., H.O.T., and J.K designed and analyzed the experiments. C.R., B.L, and S.B performed data analysis. C.R., H.O.T., and J.K wrote the manuscript with input from other authors.) Appropriate permission has been granted by the publisher.

cell fate can be divided into four phases: specification, commitment, maintenance, and differentiation (Nishioka et al., 2008; Pfeffer and Pearton, 2012). The current consensus suggests that specification of the first cell fate, which occurs prior to implantation, is activated by the Hippo signaling pathway (Nishioka et al., 2009). Nuclear localization of Yap following Hippo repression by Notch in ‘‘outer’’ cells results in the activation of critical TE gene expression through interaction with Tead4 (Rayon et al., 2014). This leads to sustained and restricted expression of Caudal-type homeobox 2 (Cdx2) and initiation of the commitment stage of the first cell fate decision. Cdx2 is broadly expressed during preimplantation development (eight- to 16-cell stage) but becomes restricted to the outer cells of the TE during blastocyst formation (Dietrich and Hiiragi, 2007). *Cdx2*$^{-/-}$ embryos develop to the blastocyst stage but fail to maintain their morphology and do not implant (Strumpf et al., 2005). Oct4 is essential for establishing embryonic pluripotency (Nichols et al., 1998; Strumpf et al., 2005) and, while broadly expressed at the same stages, becomes restricted to the ICM (Dietrich and Hiiragi, 2007). Accordingly, *Oct4*$^{-/-}$ embryos fail to establish a functional ICM (Nichols et al., 1998).

Studies in both preimplantation embryos and ES cells have established an antagonistic relationship between Cdx2 and Oct4 during TE commitment. Knockout or knockdown of Cdx2 permits expression of Oct4 in the TE lineage (Strumpf et al., 2005; Wu et al., 2010), whereas overexpression (OE) of Cdx2 or knockdown of Oct4 in ES cells induces TE differentiation (Niwa et al., 2005). Similarly, OE of Cdx2 or the additional TE-restricted TF Gata3 or Tcfap2c promotes transition of ES cells into trophoblast stem (TS)-like cells, which are similar to an *in vitro* counterpart of TE derived from preimplantation embryos (Kuckenberg et al., 2010; Ralston et al., 2010). In contrast, OE of Oct4 in TS cells promotes an ES cell-like fate (Wu et al., 2011). Several factors preferentially expressed in the TE (e.g., Cdx2, Gata3, and Tcfap2c) are involved in self-renewal of TS cells (Auman et al., 2002; Chawengsaksophak et al., 1997; Ralston et al., 2010). Although the antagonistic regulatory mechanism between Cdx2 and Oct4 has been widely accepted from results obtained from mouse ES cells, whether they

directly repress each other remains controversial (Nishiyama et al., 2009; Niwa et al., 2005).

Most TFs within the pluripotency network of ES cells are coordinately down-regulated upon exit from the self-renewal program, with only a few factors up-regulated. AT-rich interactive domain 3a (Arid3a)/Bright/Dril1 is one such pluripotency network factor whose modest expression in self-renewing ES cells is dramatically up-regulated upon differentiation (Wang et al., 2006). Arid3a, the founding member of the ARID family of TFs, has been characterized as a transactivator of both B lymphocyte development and cell cycle progression (Herrscher et al., 1995). Loss-of-function studies revealed that >98% of *Arid3a*$^{-/-}$ mice die prior to embryonic day 11.5 (E11.5) (Webb et al., 2011), suggesting a potential role in embryonic development. A recent follow-up study showed that singular loss of Arid3a is sufficient for reprogramming as well as enhancement of standard four-factor reprogramming of mouse embryonic fibroblasts (MEFs) to fully induced pluripotent stem cells (Popowski et al., 2014; Takahashi and Yamanaka, 2006). That Arid3a is expressed highly in extraembryonic trophoblast lineages that give rise to the placenta (Wu et al., 2009) led us to examine its function in ES cells and TE lineage commitment and differentiation.

Here, we present evidence that Arid3a is a critical transcriptional regulator of ES to TS-like cell *trans*-differentiation and plays an important role in the commitment and differentiation of TE rather than its specification. Induction of Arid3a in ES cells or bona fide blastula-derived TS cells triggers the TE-specific gene expression program and differentiation toward subsequent trophoblastic lineages, whereas knockdown of Arid3a compromises differentiation of ES cells. When injected into four- to eight- cell stage embryos, Arid3a OE-generated TS-like cells adopt an outside cell fate, synonymous with commitment to the TE lineage. Intersection of global gene expression, genome-wide target mapping, and cellular localization studies revealed that, upon nuclear upregulation, Arid3a acts both directly upstream of and parallel to *Cdx2* to activate key TE-specific genes while directly repressing regulators of ES cell pluripotency, including *Oct4* and

Nanog. Arid3a and histone deacetylases 1/2 (HDAC1/2) associate at the protein level and selectively co-occupy regulatory regions of pluripotent genes, suggesting a mechanism by which differential Arid3a complexes contribute to execution of the first cell fate decision.

2.2 MATERIALS AND METHODS

Cell culture and stable cell lines

Mouse J1 ES cells were maintained in DMEM (Dulbecco's modified Eagle's medium) supplemented with 18% fetal bovine serum (FBS), 2 mM L-glutamine, 100 mM nonessential amino acid, nucleoside mix (100X stock; Sigma), 100 μM β-mercaptoethanol, 1000 U/mL recombinant LIF (Chemicon), and 50 U/mL penicillin/streptomycin. ES cells were cultured in 0.1% gelatin-coated dishes. Mouse TS cells were maintained at a ratio of 3:7 of TS medium to MEF-conditioned TS medium with 25 ng/mL Fgf4 and 1 μg/mL heparin. The TS medium was RPMI 1640 (Gibco) supplemented with 20% FBS, 100 μM β-mercaptoethanol, 2mM L-glutamine, 1mMsodium pyruvate, 50 U/mL penicillin, and 50 mg/mL streptomycin. The MEF-conditioned medium was TS medium conditioned by MEF cells. Mitomycin-treated MEF cells were cultured in TS medium for 3 d. The medium was collected every 3 d for three times. 293Tcells were maintained in DMEM supplemented with 10% FBS, 2mML-glutamine, and 50 U/mL penicillin/streptomycin. All cells were incubated at 37°C and 5% CO_2.

Lentiviral production and infection

293T cells were plated at ~6 x 10^6 cells per 100-mm dish and incubated overnight. Cells were transfected with 6 μg of pLKO.1 shRNA vector (Sigma) (Appendix) with 4 μg of pCMV-Δ8.9 and 2 μg of VSVG helper plasmids using Fugene (Promega), according to the manufacturer's instructions. After 15 h, 293T medium was replaced with ES medium. Two days following transfection, the supernatants containing viral particles were collected and filtered through 0.45-μm pore size cellulose acetate filters. The cells, which

would be infected, were plated at ~1 X 10^6 cells per six-well plate with virus-containing supernatant supplemented with polybrene (Millipore).

Quantitative gene expression analysis

Total RNA was isolated from cultured cells using the RNeasy plus minikit (Qiagen). cDNA synthesis was performed with qScript cDNA supermix (Quanta). RT-qPCRs were performed using PerfeCTa SYBR Green FastMix (Quanta) with 1 μL of 20X-diluted cDNA generated from 500 ng of total RNA. RTqPCR primers were designed to amplify the junction between two exons using primer 3 (Koressaar and Remm, 2007). ChIP-qPCR primers were designed to amplify ~100-base-pair (bp) regions centered on the putative binding sites. Primer sequences for qPCR are listed in Appendix. CT values were normalized against Gapdh for gene expression and compared with Gfi1b_proximal promoter binding (negative control) for ChIP enrichment.

Microarrays and gene expression analysis

Affymetrix GeneChip mouse genome 430A 2.0 arrays were used for gene expression profiles. cDNA synthesis, labeling, hybridization, washing, and scanning were performed by the Microarray Core Facility at the Dana Farber Cancer Institutes (DFCI). Expression data were normalized using a robust multiple array (RMA).

ChIP-seq

ChIP assays were done with ES cell lines expressing BirA only (reference) or both BirA and biotin-tagged proteins (samples) as previously described (Kim et al., 2008) using streptavidin magnetic particles (Roche). ChIP-seq libraries were generated using ChIP-seq library prep kits (New England Biolabs) according to the manufacturer's instructions. ChIP-seq libraries were sequenced using an Illumina HiSeq 2500 at the Genomic Sequencing and Analysis Facility (GSAF) of The University of Texas at Austin. Raw and processed microarrays and ChIP-seq data have been deposited at the public server Gene Expression Omnibus (GEO) under accession number GSE56877.

Immunofluorescence

Glass coverslips were affixed to the bottom of six-well plates followed by 0.1% gelatin coating. The cells were plated at high density and grown for 1–3 d at 37°C. Slides were incubated in 4% paraformaldehyde (in PBS) for 15 min at room temperature followed by 5% Triton X-100 (in PBS) incubation for 10 min at room temperature. Slides were then incubated in blocking solution (3% BSA + 1% normal horse serum in PBS) for 1 h at room temperature, primary antibody solution (1:200 dilution) for 1 h at room temperature, and secondary antibody conjugated to Alexa Fluor 594 or Alexa Flour 488 dyes (1:1000 dilution) for 1 h in the dark at room temperature. Last, the glass coverslips were dried, affixed to slides using mounting solution, and imaged on a Zeiss 710 laser scanning confocal and structured illumination microscope.

Embryo chimeras

Host embryos were obtained from B6D2F1 mice and collected at the two-cell stage. Approximately 10 ES, Arid3a-overexpressing, or Cdx2-overexpressing cells expressing ZsGreen (pEF1a-IRES-ZsGreen1, Clontech) were injected into four- to eight-cell stage embryos. After injection, the embryos were cultured *in vitro* in KSOM+AA (Millipore). Images were collected at E3.5.

Generation of stable cell lines

Arid3a cDNA was amplified by PCR using a Arid3a mouse cDNA clones (Origene, MC205205) as a template and then cloned in to the pEF1α-FLBIO vector (Wang et al., 2006). Primer sequences for PCR are listed in Appendix. The Arid3a-containing vector was transfected into BirA expressing ES cells by electroporation. The positive biotin-tagged Arid3a OE stable cells were picked and maintained under the same ES media described above with puromycin (Invitrogen) and neomycin (Gibco). Ectopic expression levels were measured by RT-qPCR and Western blotting using anti-streptavidin-HRP (Invitrogen) and anti-Arid3a (Webb et al., 2011).

Alkaline phosphatase (AP) staining

AP staining was performed with an Alkaline Phosphatase Detection Kit (Millipore) according to the manufacturer's protocol. Cells were imaged using an inverted microscope (Zeiss).

Cell growth rate assay

Cell growth rate assays were performed with the Cell Counting Kit-8 (Dojindo) according to the manufacturer's protocol. The absorbance was measured at 490nm using a micro-titer plate reader.

Cell cycle analysis

Approximately $1X10^6$ cells were harvested after washing with PBS. The harvested cells were fixed in ice cold ethanol and incubated at 4°C overnight. Fixed cells were washed with PBS and then stained with propidium iodide in a solution containing RNase A. Analysis was performed on a FACSCalibur Flow Cytometer (Becton Dickinson).

Gene Ontology (GO) analysis

DAVID 6.7 (Huang da et al., 2009) was used for differentially expressed genes from expression profile data, and Genomic Regions Enrichment of Annotation Tool (GREAT) (McLean et al., 2010) was used for ChIP-seq data.

Antibodies

anti-Arid3a (1:5000;(Webb et al., 2011)), anti-Oct3/4 (1:1000, SC-5279, Santa Cruz), anti-Sox2 (1:1000, SC-17320, Santa Cruz), anti-Nanog (1:1000, ab80892, Abcam), anti-β-actin (1:10000, ab20272, Abcam), anti-HDAC1 (1:1000, ab7028, Abcam), anti-HDAC2 (1:1000, #5113, Cell Signaling), anti-Cdx2 (1:1000, CDX2-88, Biogenex), anti-Gata3 (1:1000, SC-9009, Santa Cruz), anti-Tead4 (1:1000, ab58310, Abcam), and anti-Yap1 (1:1000, SC-101199, Santa Cruz).

Western blotting

Proteins were lysed from cultured cells using RIPA Lysis and Extraction Buffer (G-Biosciences) supplemented with protease inhibitor cocktail (Roche). Cell lysates were

separated by electrophoresis on 4-20% gradient acrylamide gels and transferred to PVDF membranes. The blots were blocked with TBS-T (20mM Tris-HCl, pH 7.6, 136mM NaCl, and 0.1% Tween-20) containing 5% skim milk for an hour and then incubated with primary antibody solution at 4°C overnight. After washing with TBS-T, the membranes were incubated with HRP-conjugated secondary antibodies for 1 hr at room temperature (RT). Then the membrane was developed with ECL reagents (GE Healthcare).

Whole-mount RNA in situ hybridization

RNA in situ hybridization was carried out as previously described (Ai et al., 2007). Staining reactions were carried out for 2-72 hrs. The Arid3a antisense probe was a 390b mouse Arid3a fragment generated from a fragment spanning exons 3 and 4 cloned in pGEM-Teasy (Invitrogen) in the T7/sense orientation using primers listed in Appendix.

Whole mount embryo staining

Embryos were washed with PBS, followed by fixation in 4% PFA in PBS for 15 minutes at RT. Then, the embryos were rinsed briefly in PBS+0.2% goat serum (PBSS) at RT for three times. Next, the embryos were permeabilized with 0.2% Triton X-100 in PBS for 20 minutes at RT, followed by washing in PBSS for 5 minutes at RT. Permeabilized embryos were blocked with 0.1% BSA+2% goat serum in PBS (blocking solution) for 1hour at RT, primary antibody solution (1:200 dilution) for overnight at 4°C, and secondary antibody conjugated to Alexa Fluor 594 or Alexa Flour 488 dyes (1:1000 dilution) with DAPI (1:1000 dilution) for 1 hour in the dark at RT. Lastly, the embryos were washed three times with PBS. In the last washing, the embryos were moved to slides and imaged on a Zeiss 710 Laser Scanning Confocal & SIM microscope.

Immunohistochemistry (IHC)

Preparation and fixation of mouse placentas (C57BL/6) at E11.5 were carried out as previously described (Joyner and Wall 2008). Sectioning was done at histology core, the University of Texas at MD Anderson Smithville. Antibodies used for staining were anti-

Arid3a (1:5000;(Webb et al., 2011)), anti-Proliferin (SC-47347, Santa Cruz), and anti-Tpbpa (ab104401, Abcam).

Animals

All housing, husbandry, and experimental procedures with Arid3a knockout and control mice were approved by the Institutional Animal Care and Use Committees at The University of Texas at Austin. The generation of Arid3a$^{-/-}$ mice has been previously described (Webb et al., 2011). Mice used in for placental collection and IHC were PCR genotyped using embryo tail genomic DNA.

Immunoprecipitation (IP)

One-step affinity purification with streptavidin-agarose bead (Invitrogen) using ~2 x 10^7 cells from ES cell lines expressing BirA only or BirA with biotin-tagged proteins was performed as previously described (Kim et al., 2010). The final pulled-down proteins were eluted in Laemmli buffer prior to Western blotting.

Principal component analysis (PCA)

PCA analysis was performed using published expression profiles of TS and differentiated TS cells (Kidder and Palmer, 2010) with those of our Arid3a OE cell lines. Common probe sets from different Affymetrix array platforms were extracted, followed by normalization of the probe sets using Robust Multi-array Average (RMA) analysis from Bioconductor. PCA was performed using excel add-ins PCA analysis program, multibase 2015 (http://www.numericaldynamics.com/).

Identification of ChIP-sequencing peaks

FASTAQ files were aligned to mouse genome (mm9, NCBI Build 37) using a Burruws-Wheeler Aligner (BWA) allowing up to two mismatches (Li and Durbin, 2009). For the identification of transcription factor (TF) binding sites, model-based analysis with the ChIP-seq (MACS) peak caller (Zhang et al., 2008) was used with a default setting. The ChIP-seq signals resulting from BirA-only expressing ES cells were subtracted to remove background noise for Arid3a OE cells. Genes occupied by TFs within a 10Kb window

(from 8Kb upstream to 2Kb downstream of TSS were considered as target genes and used for subsequent analyses.

Correlation map

Common binding sites of TFs indicated in Figure 2.14A were identified by peak calling with an overlap analysis. The binding sites of two TFs were calculated for with a paired-end Pearson correlation coefficient.

Motif analysis

Motif searches were performed with motif-based sequence analysis tools (Bailey et al., 2009). Motif similarities were calculated using TOMTOM (Gupta et al., 2007).

Public data sets used

Gene expression data sets used for the analyses shown in Figure 2.5,2.15G were obtained from the Gene Expression Omnibus (GEO) database under the accession numbers of GSE 12986 (Cdx2 OE and Gata3 OE cells), GSE 3766 (TS cells), GSE 2204 (XEN cells), GSE 3729 (ES cells), GSE 18507 (TS cell differentiation), and GSE20177 (TS and ES cells). ChIP-sequencing data sets used for the analyses shown in 13 were obtained from GSE 11431 (E2F1 and n-Myc), GSE 11724 (Nanog and Sox2), GSE 44288 (Oct4), GSE 13084 (Ezh2), GSE 26680 (Ring1b), GSE 18776 (Suz12), and GSE 16375 (Cdx2).

2.3 RESULTS

Arid3a is dispensable for mouse ES cell self-renewal but critical for normal differentiation

In order to elucidate the function of Arid3a in self-renewal and differentiation, we first performed a shRNA-based knockdown of Arid3a in the mouse ES cell line J1. We obtained >75% knockdown efficiency at both the protein and mRNA levels (Figure 2.1A,B). Arid3a-deficient cells displayed normal ES cell morphology as well as alkaline phosphatase (AP) activity and proliferation rates comparable with control cells (Figure 2.1C,D). While knockdown of Arid3a resulted in modest up-regulation of mRNA levels of previously established ES cell core pluripotency factors, including Oct4, Sox2, and

Nanog (Figure 2.1B), we observed no significant changes in the protein levels of these factors (Figure 2.1A). These data indicate that Arid3a is dispensable for self-renewal of ES cells.

Unlike the majority of TFs comprising the ES cell pluripotency network, Arid3a is moderately expressed in undifferentiated ES cells but significantly induced upon differentiation (Figure 2.2A). To test the potential function of Arid3a during differentiation, we cultured Arid3a knockdown ES cells and normal ES cells in differentiation medium (leukemia inhibitory factor [LIF] withdrawal). While control ES

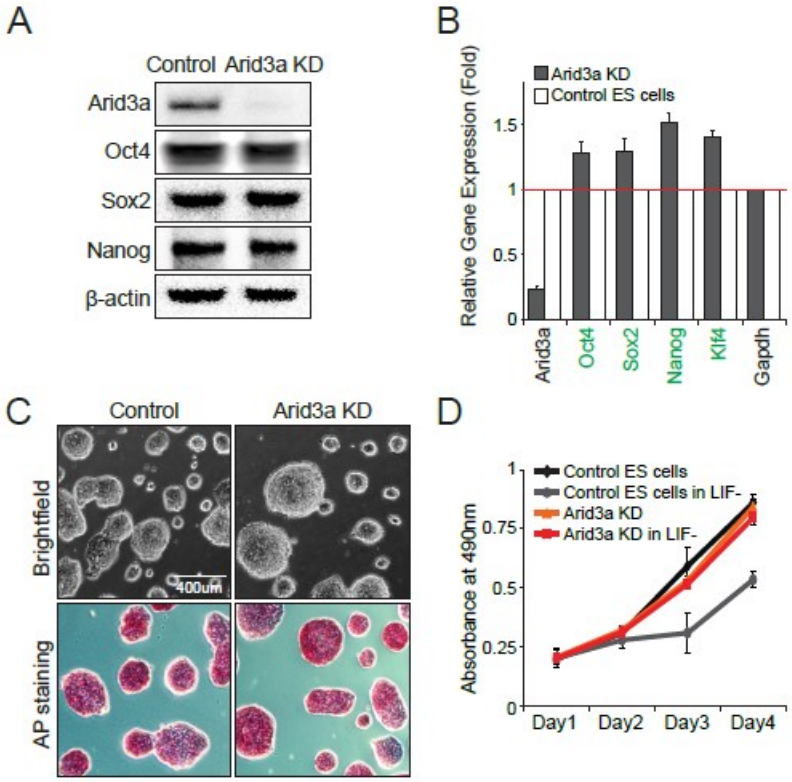


Figure 2.1 Arid3a is dispensable for self-renewal of ES cells. (A) Protein levels of the pluripotency-associated TFs, Oct4, Sox2, and Nanog, following Arid3a KD as measured by Western blotting. β-actin is the loading control. (B) Arid3a KD (gray) in ES cells leads to modest induction of pluripotency-associated TF expression as determined by RT-qPCR. The data are plotted relative to corresponding expression levels of control ES cells (white). (C) ES cell morphology upon KD of Arid3a. Arid3a KD cells were plated at clonal density, cultured for 5 days, and stained for alkaline phosphatase (AP). (D) Comparison of growth rates of Arid3a KD cells with normal ES cells under LIF+ and LIF- conditions.

cells showed differentiated morphology with reduced rates of growth, Arid3a-deficient ES cells maintained undifferentiated cell morphology for about five additional days, with proliferation rates similar to those of control ES cells under normal culture conditions with LIF (Figure 2.1D,2.2B). The results suggested that either Arid3adeficient ES cells are less dependent on LIF, or the cells have impaired differentiation potential. However, after prolonged culture in the absence of LIF, we began to observe differentiated morphology of Arid3a-deficient cells (Figure 2.2B), suggesting that its loss delayed normal ES cell differentiation. Consistent with this, we observed delayed induction of lineage-specific marker genes such as Brachyury (mesoderm), Gata6 and Sox17 (endoderm), and Cdx2 and Gata3 (TE) as well as delayed down-regulation of ES cell core factors (Figure 2.2C,D). Notably, the most significantly affected genes upon knockdown of Arid3a were TE markers (Figure 2.2C), suggesting that Arid3a may play a role in TE differentiation. We confirmed these results in an independent ES cell line (E14) and by using a shRNA targeted to the Arid3a 3' untranslated region (UTR) (Figure 2.2D-H). The delayed differentiation observed upon knockdown of Arid3a was successfully rescued by OE of shRNA-resistant Arid3a (Figure 2.2G,H).

Induction of Arid3a in ES cells promotes TE differentiation

The results from knockdown of Arid3a suggested that the reciprocal action, OE of Arid3a, might promote ES cells to differentiate even in the presence of LIF. To examine this possibility, we generated a pool of Arid3a-overxpressing ES cells (Arid3a OE-Pool) and, from this, multiple clones with different levels of ectopic Arid3a expression (representative clones are shown in Figure 2.3A,B). Whereas clones expressing Arid3a modestly above endogenous levels (e.g., Arid3a OE-Low; approximately two-fold) displayed normal ES cell morphology (Figure 2.3A,B), clones with ~20-fold OE (e.g., Arid3a OE-High) displayed primarily differentiated phenotypes (flattened epithelial-like morphology) with weak AP activity (Figure 2.3C,D). Arid3a OE-High cells exhibited significantly slower proliferative rates than normal ES cells (Figure 2.3E). Subsequent cell cycle analysis confirmed that Arid3a OE-High cells were reduced in actively

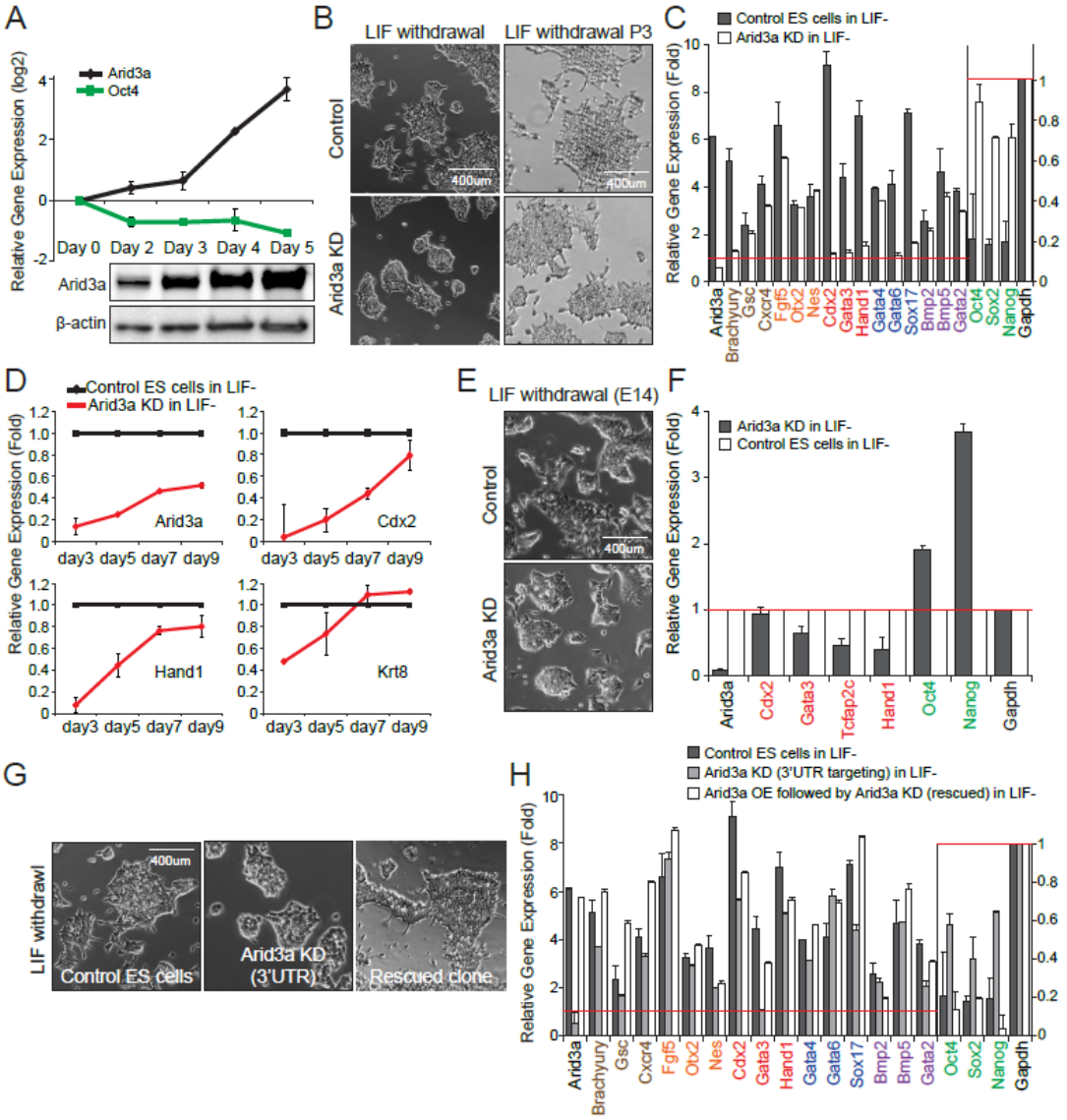


Figure 2.2 Arid3a is critical in ES cell differentiation. (A) Reciprocal expression levels of Arid3a (black) and Oct4 (green) upon differentiation of ES cells as determined by RT-qPCR (top panel) and Western blotting. (B) ES cell morphologies of control ES cells and Arid3a KD cells upon LIF withdrawal. (C) Expression analysis of lineages and pluripotency-associated genes (brown, mesoendoderm; orange, ectoderm; red, trophectoderm; blue, endoderm; purple, mesoderm; and green, pluripotency-associated) by RT-qPCR in control (gray) and Arid3a KD ES cells (white) following removal of LIF. (D) Up-regulation of expression levels of TE lineage-specific transcripts is delayed in Arid3a KD cells (red) relative to control ES cells (black) following removal of LIF. (E) Cell morphologies of control ES cells and Arid3a KD cells upon LIF withdrawal, passage 2. (F) Expression analysis of lineages and pluripotency-associated genes in control (white) and Arid3a KD ES cells (gray) following removal of LIF. (G) Comparison of cell morphologies of control ES, Arid3a KD (targeting 3'UTR), and Arid3a OE followed by Arid3a KD (rescued clone) cells upon LIF withdrawal, passage 2. (H) Expression analysis of lineage- and pluripotency-associated genes by RT-qPCR in control ES (dark gray), Arid3a KD ES (light gray), and Arid3a OE followed by Arid3a KD (rescued, white) cells upon removal of LIF, relative to normal ES cells. Error bars depict standard deviations of biological triplicates.

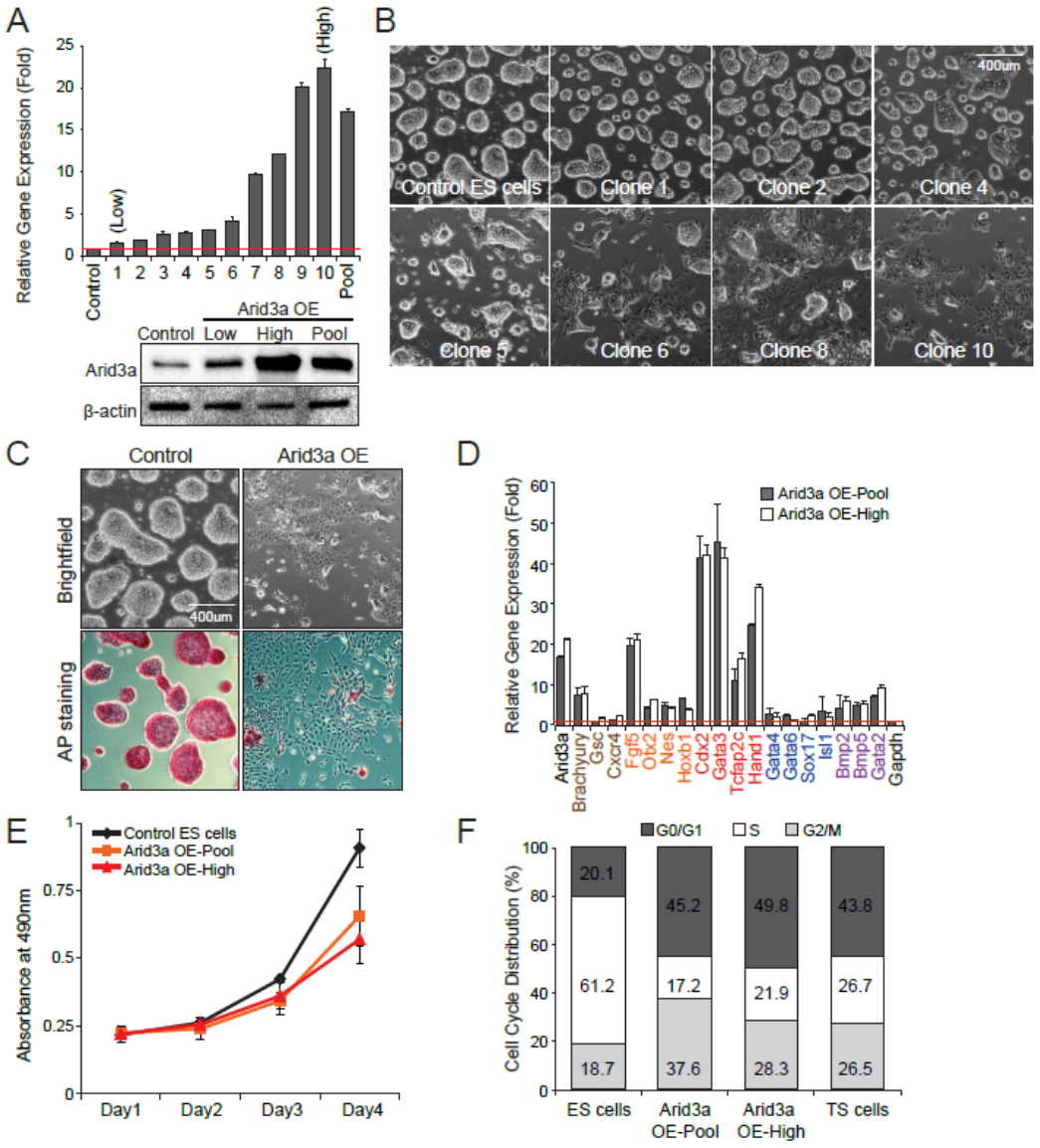


Figure 2.3 Induction of Arid3a in ES cells promotes TE differentiation. (A) Expression levels of Arid3a as determined by RT-qPCR in ten different Arid3a overexpressing (OE) clones and Arid3a OE-Pool. Clones expressing different levels of Arid3a are indicated as different numbers, ranked from 1 to 10 (Low to High). The data are plotted relative to corresponding expression levels of Arid3a in control ES cells (top panel). Protein levels of Arid3a in representative Arid3a OE cells as measured by Western blotting. β-actin is the loading control (bottom panel). (B) Cell morphologies of Arid3a OE ES cell clones expressing different levels of Arid3a. (C) ES cell morphology upon OE of Arid3a. Arid3a OE cells were plated at clonal density, cultured for 5 days, and stained for AP. (D) Expression analysis of lineage markers by RT-qPCR in Arid3a OE-Pool (gray) and -High (white) cells. The data are plotted relative to corresponding expression levels of control ES cells. All error bars depict standard deviations of biological triplicates. (E) Arid3a OE cells grow slower than control ES cells. Shown are the averages of biological triplicates with standard deviations. (F) Cell cycle comparisons of Arid3a OE cell lines (Arid3a OE-Pool and -High) with control ES and TS cells. Y-axis, distribution of cell cycle as measured by flow cytometry/DNA content.

proliferating (S-phase) populations and increased in G2/M-phase populations, which is similar to that of TS cells (Figure 2.3F). Expression of a panel of multi-lineage markers revealed that many were activated by OE of Arid3a in ES cells, whereas pluripotency-associated gene levels were reduced (Figure 2.3D,2.4A-C). TE lineage markers such as Cdx2, Gata3, Tcfap2c, and Hand1 showed particularly strong up-regulation upon OE of Arid3a in both J1 and E14 ES cells (Figure 2.3D,2.4D-F), further indicating a role for Arid3a in TE lineage specification.

Global expression profiles of Arid3a-overexpressing ES cells are similar to those of TS cells

To better understand the global transcriptional impact of Arid3a perturbation in ES cells, we compared expression profiles of Arid3a knockdown and Arid3a OE-High cells with control ES cells using Affymetrix microarrays. We identified 1842 differentially expressed genes, with a cut-off threshold of 1.5-fold in at least one expression profile of Arid3a-overexpressing cell lines (Arid3a OE-High and OE-Pool). As expected from the data shown in Figure 2.1 and 2.2, gene expression changes upon knockdown of Arid3a were not significant compared with the changes upon OE of Arid3a (Figure 2.5A).

A prior study revealed that ~50% reduction of Oct4, a master regulator of pluripotency, is sufficient to convert ES cells to TS-like cells (Niwa et al., 2000) that share many features with authentic TS cells derived from the TE layer of the early stage blastocyst. Since OE of Arid3a in ES cells significantly up-regulated many of the same TE lineage-specific markers (Figure 2.3D) as knockdown of Oct4, we also compared the global expression profiles of Arid3a-overexpressing cell lines with those of TS-like cells derived by knockdown of Oct4 in ES cells. We observed highly similar gene expression patterns for Arid3a-overexpressing and Oct4 knockdown cells (Figure 2.5A). Consistent with the propensity of Arid3a OE to promote TE lineage specification, the up-regulated genes upon OE of Arid3a were strongly enriched in gene ontology (GO) terms associated with placental and early embryonic development (Figure 2.5B). On the other hand, the relatively fewer number of genes down-regulated upon OE of Arid3a were enriched in

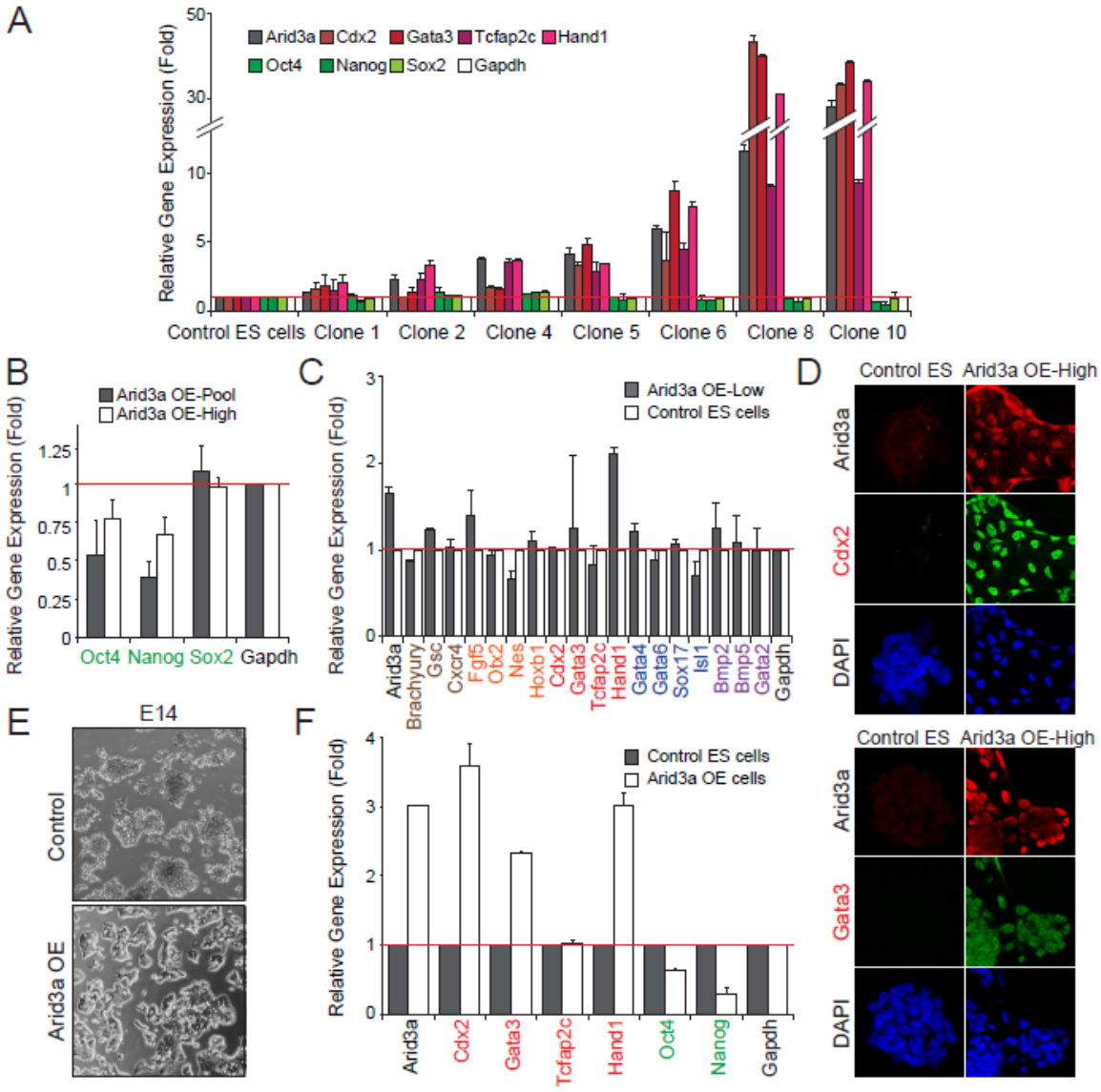


Figure 2.4. Arid3a plays a role in TE lineage specification. (A) Expression analysis of TE (red) and pluripotency-associated (green) genes by RT-qPCR in Arid3a OE cell lines from Figure 2.3B. (B) OE of Arid3a reduces levels of Oct4 and Nanog relative to control ES cells as determined by RT-qPCR. Arid3a OE-Pool (gray) and -High (white) cells data are plotted relative to corresponding expression levels of control ES cells (red horizontal line). (C) Expression analysis of lineages and pluripotency-associated genes (brown, mesoendoderm; orange, ectoderm; red, trophectoderm; blue, endoderm; purple, mesoderm; and green, pluripotency-associated) by RT-qPCR in Arid3a OE-Low (gray) and control ES (white) cells. (D) Arid3a OE-High cells express more Cdx2 (top) and Gata3 (bottom) relative to control ES cells as determined by immunofluorescence (IF). Cells were stained with anti-Arid3a (red, Alex 594), anti-Cdx2 (green, Alexa 488), anti-Gata3 (green, Alexa 488), and DAPI (blue). (E) Cell morphologies of control and Arid3a OE in E14 ES cells. OE of Arid3a in ES cells was done by transient transfection. (F) Expression analysis of TE (red) and pluripotency-associated (green) genes by RT-qPCR in Arid3a OE (gray) and control (white) in E14 ES cells. Error bars depict standard deviations of biological triplicates.

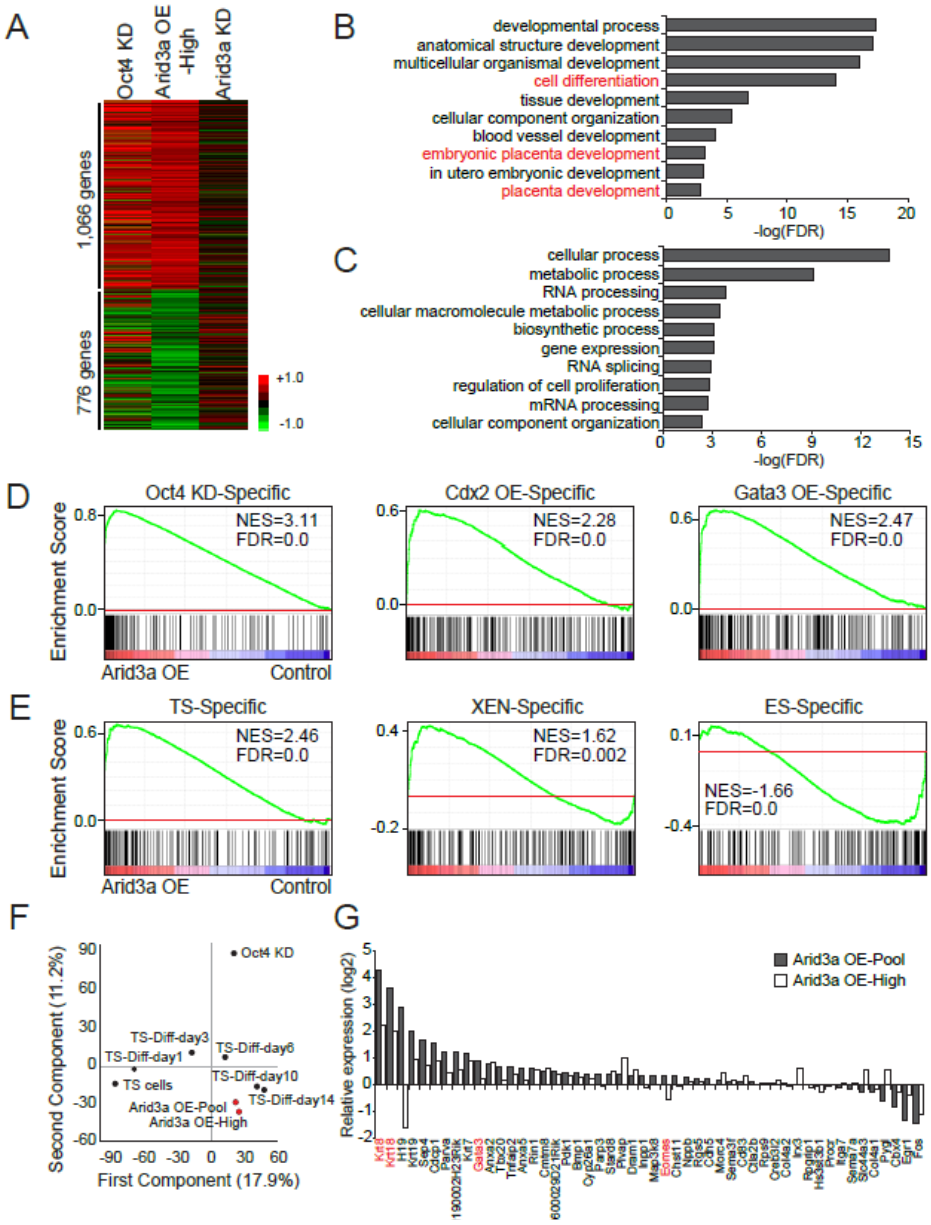


Figure 2.5 Expression profiles of Arid3a OE ES cells are highly similar to those of TS-like cells. (A) An unsupervised hierarchical clustering of expression profile of Arid3a OE-High cells. (B,C) Significantly enriched terms (biological functions) of up- (B) and down-regulated (C) genes upon OE of Arid3a shown in Figure 2.5A by Gene Ontology (GO) analysis. (D,E) Gene set enrichment analysis (GSEA) using ordered gene expression levels from Arid3a OE cells over control ES cells (X-axis) with gene sets indicated; top 1% of up-regulated genes upon Oct4 KD (204 genes), Cdx2 OE (368 genes), and Gata3 OE (384 genes) in ES cells (D). GSEA using TS cell-specific (313 genes), XEN cell-specific (227 genes), and ES cell-specific (218 genes) gene sets (E). (F) Global expression profiles of Arid3a OE cell lines (Arid3a OE-High and -Pool), TS and differentiated TS cells as determined by principal component analysis (PCA). (G) Relative expression of top 50 TS cell-specific genes in Arid3a OE-Pool (gray) and -High (white) ES cells. ES cells are used as a control. Gene names in red are well-known TE markers.

unrelated terms (e.g., RNA processing and cellular metabolic processes) (Figure 2.5C), further indicating that the cells are indeed differentiated, particularly toward the TE lineage.

Since ES cells are derived from the ICM of blastocysts, they are generally thought to have lost or suppressed the ability to generate TE and trophoblast lineages (Reubinoff et al., 2000). Nonetheless, several studies have reported that ES cells can directly *trans*-differentiate toward a TS-like cell fate (Kuckenberg et al., 2010; Niwa et al., 2005; Ralston et al., 2010). Therefore, we performed gene set enrichment analysis (GSEA) to further compare gene expression profiles of Arid3a-overexpressing cells with those of published TS-like cells generated by OE of Cdx2 or Gata3 (Ralston et al., 2010) or knockdown of Oct4 in ES cells. As shown in Figure 2.5D, each of the three gene sets (top 1% of genes up-regulated within each of the aforementioned profiles) showed highly significant enrichment in genes up-regulated upon OE of Arid3a in ES cells.

We carried out further GSEA with gene sets generated from the additional published expression data sets: TS cell-specific (TS cells vs. ES cells) (Ralston et al., 2010), extraembryonic endoderm (XEN) cell-specific (XEN cells vs. ES cells) (Ralston et al., 2010), and ES cell-specific (ES cells vs. differentiated ES cells) (Hailesellasse Sene et al., 2007). Consistent with our findings shown in Figure 2.3 and 2.5D, only TS cell-specific genes were strongly enriched among the up-regulated genes following OE of Arid3a in ES cells (Figure 2.5E). While XEN cell-specific genes showed marginal enrichment, as predicted, ES cell-specific genes showed a negative correlation (Figure 2.5E). Principal component analysis (PCA) indicated that Arid3a-overexpressing cells fall into a similar component with differentiated TS cells (days 10 and 14) (Figure. 5F)(Kidder and Palmer, 2010). Furthermore, the expression levels of the top 50 most strongly expressed genes in TS cells strongly overlapped with those in Arid3a-overexpressing ES cells (Figure 2.5G). Collectively, the findings indicate that OE of Arid3a in ES cells results in loss of ES cell identity and activation of TS cell-like global gene expression programs.

Induction of Arid3a in ES cells results in stable TS-like cells that are capable of further differentiation to trophoblast lineages

Although prior reports showed that OE of TE marker genes (e.g., Cdx2 and Gata3) in ES cells induces TE specification, not all of them are capable of producing stable, self-renewing TS-like cell lines (e.g., Gata3) (Niwa et al., 2005; Ralston et al., 2010). To examine whether induction of Arid3a in ES cells can lead to the establishment of stable and self-renewing TS-like cells, we adopted Arid3a OE-High cells to an established TS medium containing Fgf4 and heparin (Tanaka et al., 1998). After four passages, the cells displayed TS cell morphology; i.e., the majority of AP-negative cells gave rise to large and flattened cuboidal cells with distinctive boundaries (Figure 2.6A). Moreover, these cells expressed multiple TE marker genes at levels similar to their counterparts in authentic TS cells derived from the TE of the blastocysts (Figure 2.6B).

To test the potential of Arid3a-overexpressing cells to drive further TE differentiation, Arid3a-overexpressing TS-like cells were induced to differentiate by depleting Fgf4 and heparin from the TS culture medium. Within four to five passages, the cells appeared largely differentiated, with numerous giant cells present throughout the culture (Figure 2.6C). These cells expressed markers for more specialized cells of the TE lineage, including Hand1, Id2, Dlx3, Wnt2, and Tpbpa (Figure 2.6D), suggesting that Arid3a OE is not only sufficient for establishing the TE-like state but also may be required for further differentiation of the TE lineage.

Both Arid3a and Cdx2 are highly expressed in placenta and extraembryonic tissue such as TE (Figure. 6E)(Wheeler et al., 2003). However, analysis of public gene expression profiles (Kidder and Palmer, 2010) indicated that Arid3a, but not Cdx2, is highly up-regulated in differentiated TS cells (Figure 2.6F). This observation prompted us to examine the effect of Arid3a OE on further differentiation of TS cells. Unlike control TS cells, TS cells with OE of Arid3a more closely resembled trophoblast giant cells (TGCs) even in the presence of Fgf4 and heparin (Figure 2.6G). Accordingly, we observed up-regulation of transcripts (Gcm1, Hand1, Id2, Dlx3, Wnt2, and Tpbpa) preferentially expressed in TGCs and other more specialized trophoblastic lineages

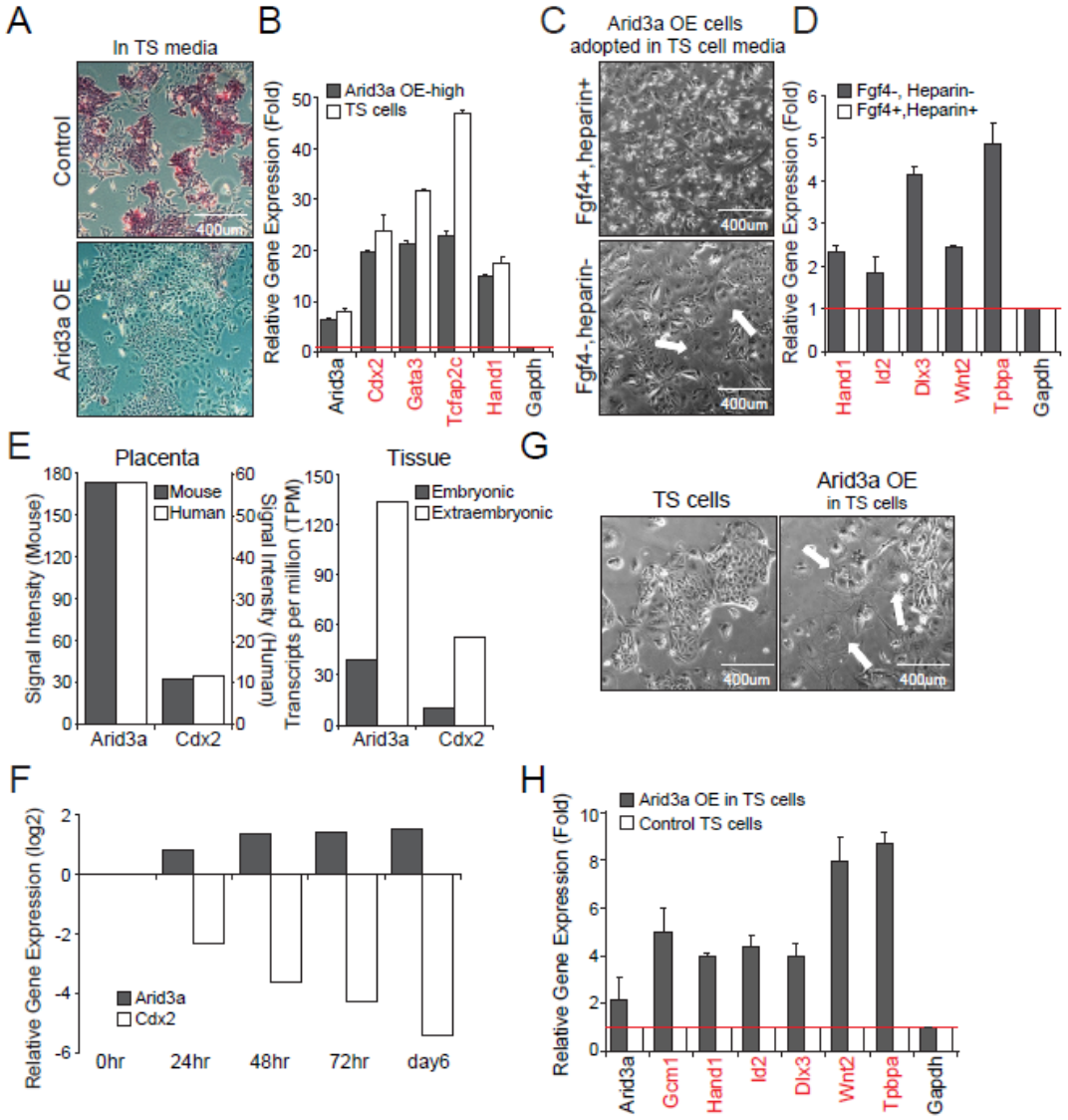


Figure 2.6. Induction of Arid3a converts ES cells to stable TS-like cells that can further differentiate to trophoblast lineages. (A) Cell morphologies of control ES and Arid3a OE cells in TS cell derivation media. (B) Expression analysis of TE markers (red) by RT-qPCR in Arid3a OE cells in TS cell derivation media (gray) and TS cells (white). (C) Cell morphologies of Arid3a OE cell adapted to TS cell derivation media (top) followed by removal of Fgf4 and heparin (bottom). Arrows indicated trophoblast giant cells (TGCs). (D) Expression analysis of TE and trophoblast markers (red) as determined by RT-qPCR in Arid3a OE cells adapted to TS cell media upon withdrawal of Fgf4 and heparin (gray). (E) Left, Arid3a and Cdx2 placental transcript levels and right, embryonic and extraembryonic transcript levels in mouse (white) and human (gray). Adapted from publicly available data sets in BioGPS (Wu et al. 2009) and UniGene (Wheeler et al. 2003), respectively. (F) Relative expression levels of Arid3a (gray) and Cdx2 (white) under TS cell differentiation culture conditions normalized to their expression levels in TS cells prior to differentiation. (G) Cell morphologies of control TS and Arid3a OE TS cells under TS cell derivation conditions. Arrows indicate TGCs. (H) Expression analysis of TE differentiation markers (red) upon OE of Arid3a in TS cells (gray) and in control TS cells (white) by RT-qPCR. Error bars depict standard deviations of biological triplicates.

(Figure 2.6H). Thus, in contrast to Cdx2, whose function is restricted to TE lineage specification, Arid3a regulates additional trophoblastic pathways beyond TE specification, as its OE in TS cells is sufficient to drive further trophoblastic differentiation.

Arid3a and Cdx2 act both independently and in concert to antagonize Oct4 and promote ES cells to TS cell *trans*-differentiation

To address whether Arid3a participates in the previously proposed Cdx2–Oct4 regulatory loop, we compared the consequences of Oct4 knockdown with double knockdown of Oct4 + Arid3a or Oct4 + Cdx2 in normal ES cell culture medium. As expected, control Oct4 knockdown cells displayed differentiated morphology (Figure 2.7A) with strong up-regulation of TE markers, including both Cdx2 and Arid3a (data not shown). As with Oct4 + Cdx2 double knockdown, double knockdown of Oct4 + Arid3a led to significant reduction of the differentiated morphology with less induction of TE markers than control Oct4 knockdown cells (Figure 2.7A,B). Immunostaining for Cdx2, Oct4, and Nanog confirmed these results (Figure 2.7C), indicating that Arid3a is at least in part involved in the Cdx2–Oct4 antagonistic regulatory loop.

To test whether the involvement of Arid3a in the Cdx2–Oct4 loop is through Cdx2 or independent of Cdx2, we performed Arid3a OE in combination with Cdx2 knockdown as well as Cdx2 OE in Arid3a knockdown cells. Cdx2 knockdown + Arid3a-overexpressing cells and Arid3a knockdown + Cdx2-overexpressing cells both showed differentiated morphology with induction of TE markers and repression of pluripotency genes (Figure 2.7D,E). Thus, Arid3a specifies TE not only through Cdx2 but also independent of Cdx2 and vice versa. That is, Arid3a and Cdx2 can induce TE-like gene expression programs in the absence of each other, although the effect is stronger when both are co-expressed.

Arid3a gain of function promotes incorporation of ES cells into the TE of developing embryos, whereas its loss of function leads to defective placentas To examine the potential of Arid3a induction to establish the outside cell fate synonymous with *in vivo* commitment to the TE lineage, we marked Arid3a OE-driven TS-like cells with ZsGreen

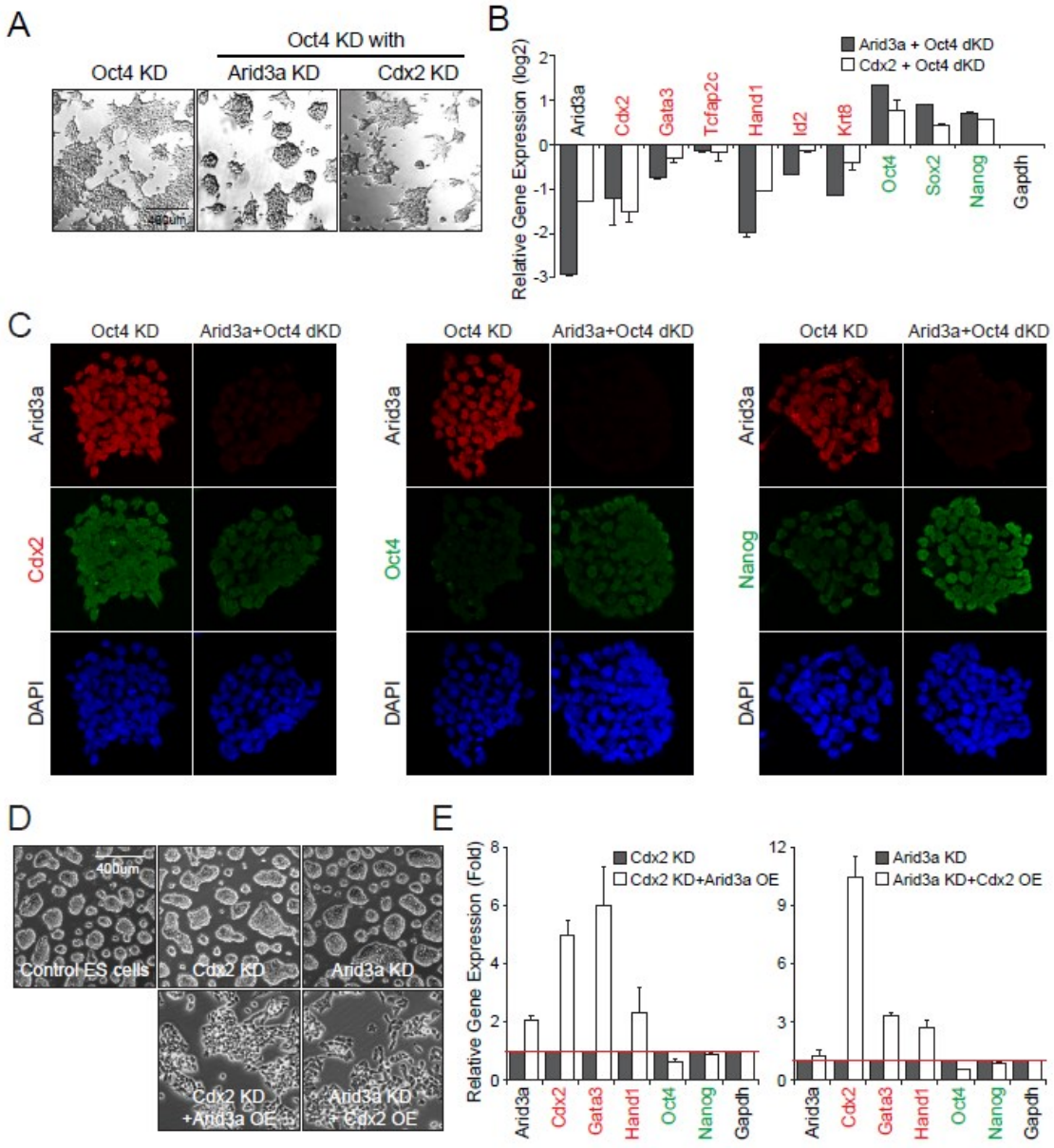


Figure 2.7. Induction of Arid3a in ES cells gives rise to TS-like cells capable of further differentiation to trophoblast lineages. (A) Cell morphologies upon KD of Oct4 and double KD of Oct4+Arid3a or Oct4+Cdx2 in ES cells under normal ES culture conditions. (B) Expressions of TE (red) and pluripotency-associated (green) genes as determined by RT-qPCR following the indicated KD perturbations (KD or double KD; dKD) in ES cells (LIF+ conditions). The data are plotted relative to corresponding expression levels in Oct4 KD cells. (C) IF images of Oct4 KD and Arid3a+Oct4 double KD cells. Cells were stained with anti-Arid3a (red, Alexa 594), anti-Cdx2, anti-Oct4, or anti-Nanog (green, Alexa 488), and DAPI (blue). (D) Both Cdx2 KD+Arid3a OE cells and Arid3a KD+Cdx2 OE cells display differentiated morphology compared to control ES, Cdx2 KD, and Arid3a KD cells. (E) Cdx2 KD+Arid3a OE and Arid3a KD+Cdx2 OE cells show increased levels of TE markers (red) and decreased levels of pluripotency-associated genes (green) relative to those of Cdx2 KD and Arid3a KD cells, respectively.

(Figure 2.8A) and injected them into four- to eight-cell stage mouse embryos, with Cdx2-ZsGreen-overexpressing TS-like cells as a positive control (Lu et al., 2008). At E3.5, control ES cells localized within the ICM, whereas Arid3a- and Cdx2-overexpressing cells were found exclusively within the TE layer (Figure 2.8B). *Arid3a*$^{-/-}$ null (knockout)-derived ES cells (Webb et al., 2011) showed normal ES cell morphology (Figure 2.8C) but also failed to localize to the outer TE layer (Figure 2.8B). These along with the above *in vitro* data indicate that gain of function of Arid3a is sufficient to not only direct TE fate but induce *trans*-differentiation of ES cells into TS-like cells with homing properties morphologically indistinguishable from bona fide TS cells.

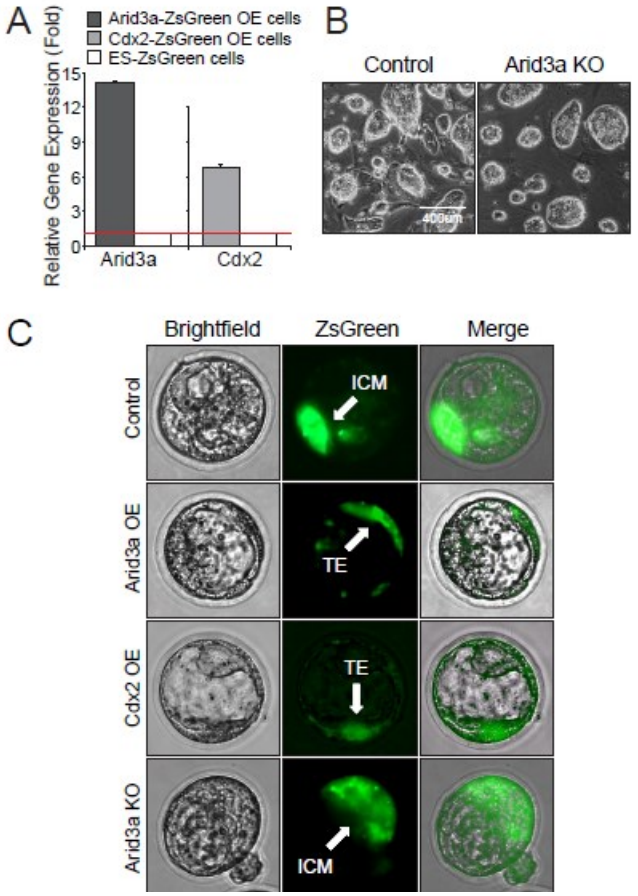


Figure 2.8. Arid3a OE promotes incorporation of ES cells into TE of developing embryos. A) Expression levels of Arid3a (gray) and Cdx2 (light gray) as determined by RT-qPCR in Arid3a OE (left) and Cdx2 OE (right) cells that were used for the embryo injections of Figure 2.8C. Error bars depict standard deviations of technical triplicates. (B) Cell morphologies of control ES and Arid3a KO cells. (C) Indicative cells were injected into 4- to 8-cell stages of embryos. ZsGreen-Arid3a OE TS-like cells integrate into the TE layer of developing embryos.

While the cause of death of *Arid3a*-null embryos was assigned to failed erythropoiesis at E11.5 (Webb et al., 2011), potential placental defects were not examined. Genotypes of earlier developmental time points revealed loss of knockout Mendelian ratios as early as E8.5 (Figure 2.9A). In situ hybridization of *Arid3a*$^{+/+}$ embryos at E6.5 indicated strong Arid3a expression in the ectoplacental cone and extraembryonic ectoderm of the chorion—sites at which multipotent TS cells reside (Figure 2.9B)(Uy et al., 2002). Placental cell types that derive from these regions—TGCs and spongiotrophoblasts (SpTs)—strongly expressed Arid3a within their nuclei, as shown in E11.5 *Arid3a*$^{+/+}$ sections by immunohistochemistry (IHC) (Figure 2.9C,D). IHC of E11.5 sections of *Arid3a*$^{-/-}$ placentas revealed multiple abnormalities. These included (1) reduction and disorganization of TGCs and the TGC cell layer, with multiple TGCs aberrantly located in the SpT and labyrinth layers; (2) lack of an organized and compact SpT layer; (3) reduction in the number of fetal blood vessels in the labyrinth (particularly in the central region); and (4) reduction in the number of maternal blood spaces in the labyrinth (Figure 2.9D). Thus, Arid3a is a critical regulator for not only TE lineage maintenance and differentiation but also proper placenta development.

Nuclear localization of Arid3a is required for Oct4 repression and TE differentiation

That Arid3a undergoes CRM1-dependent nucleocytoplasmic shuttling in B lymphocytes and other somatic cell types (Kim and Tucker, 2006) led us to test its localization in the blastocyst and prior to the first segregation by whole-mount embryo staining. In the compacted morula stage, Arid3a is expressed in all blastomeres, with localization primarily in the cytoplasm (Figure 2.10A). At E.3.5 (early blastocyst), Arid3a localizes only within the cytoplasm of the ICM but begins to accumulate within the nucleus of the outer cells (Figure 2.10B). Around E.4.5 (late blastocyst stage), Arid3a is highly localized in the outer cells of the TE (Figure 2.10A).

We next examined Arid3a localization relative to Oct4 in normal ES cells, Arid3a-overexpressing cell lines, and TS cells by immunofluorescence. As shown in Figure 10C,

A

	Arid3a+/+	Arid3a+/-	Arid3a-/-	Total
E8.5	22 (22)	50 (43)	14 (22)	86
E9.5	49 (42)	101 (84)	18/4 (42)	168
E10.5	24/1 (29)	75/1 (58)	17/1 (29)	116
E11.5	81/6 (63)	138 (126)	33/7 (63)	252
E12.5	90 (50)	115 (100)	5/3 (50)	200
E13.5	83 (61)	160 (122)	0 (61)	243

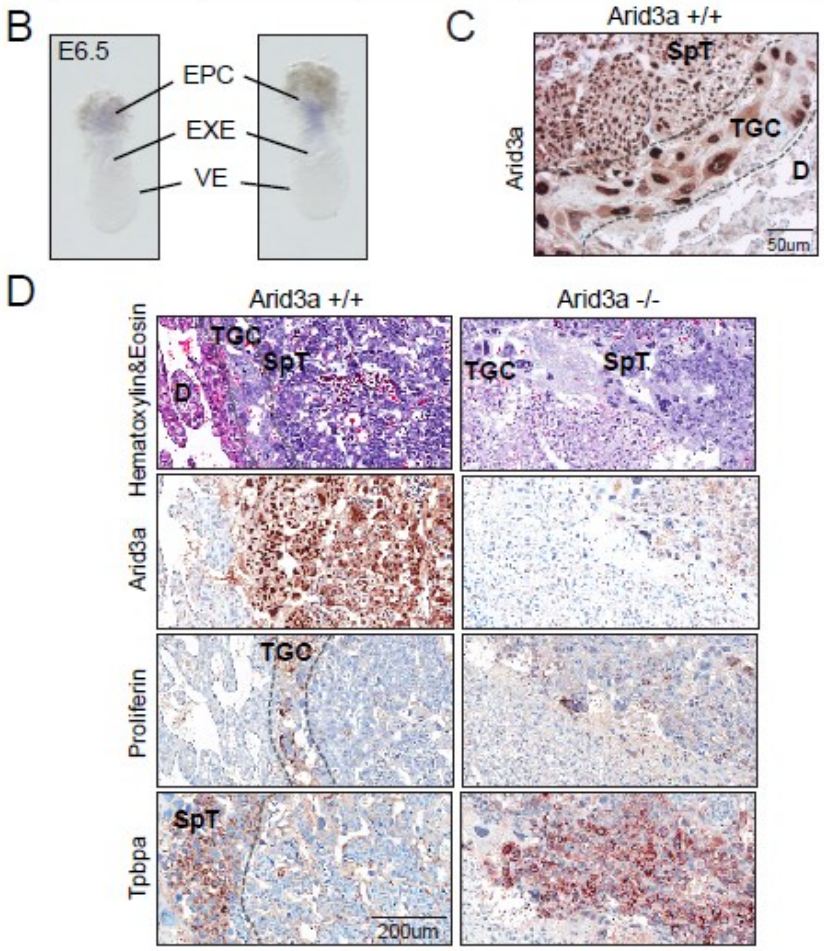


Figure 2.9 Loss of Arid3a leads to defective placentas. (A) Genotypes of offspring from heterozygous matings of Arid3a+/- mice. Mendelian frequencies, parentheses; absorptions, slashes. (B) In situ hybridization of Arid3+/+ embryos at E6.5 showing Arid3a expression in the ectoplacental cone and chorion. EPC, ectoplacental cone; EXE, extraembryonic ectoderm; VE, visceral endoderm. (C) Arid3a is strongly expressed within nuclei of trophoblast giant cells (TGCs) and spongiotrophoblasts (SpTs) of E11.5 Arid3a+/+ placentas. (D) Placentas of E11.5 Arid3a -/- embryos are defective. IHC was performed with anti-Proliferin, which marks trophoblast giant cells (TGCs), and anti-Tpbpa, which stains Spongiotrophoblasts (SpTs); D, Deciduum. Anti-Arid3a IHC revealed high levels of Arid3a expression in Arid3a+/+ TGCs and SpTs. The dotted lines denote the wildtype TGC and SpT expression domains that are grossly disorganized in the null placentas.

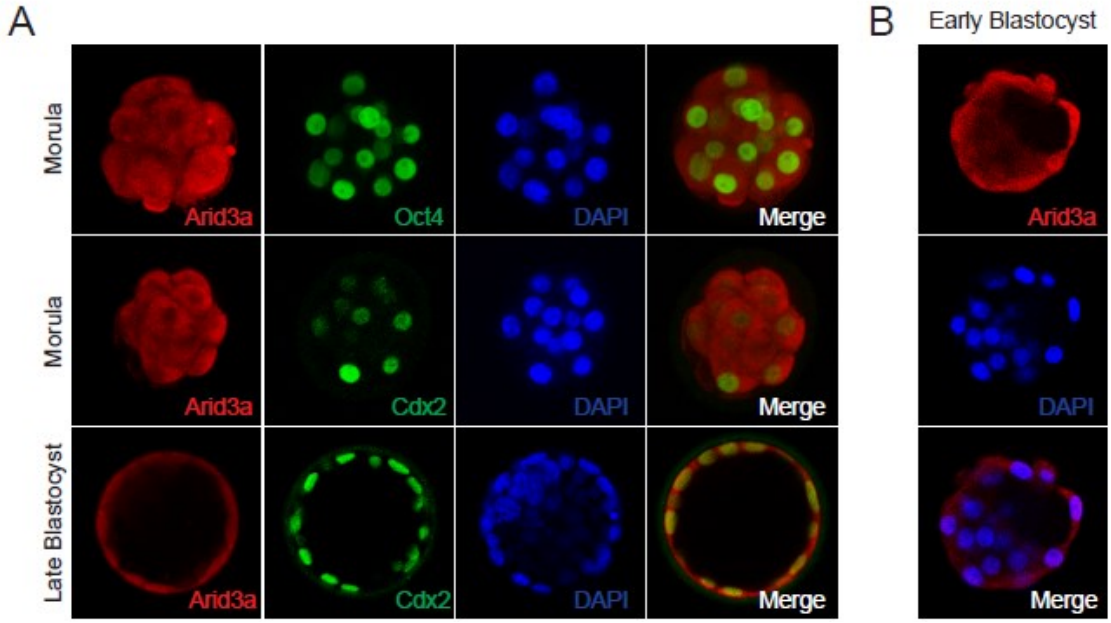


Figure 2.10. Arid3a is localized in TE of developing embryo. (A-B) Whole mount embryo staining of compacted morula, late blastocyst (A), and early blastocyst stages (B) with anti-Arid3a (red, Alexa 594), anti-Oct4 (green, Alexa 488), anti-Cdx2 (green, Alexa 488), and DAPI (blue). Arid3a localizes within all blastomeres of the compacted morula and primarily within the outer cells of TE in blastocyst.

Arid3a, in contrast to Oct4, localized primarily within the cytoplasm of ES and Arid3a OE-Low cells. However, the significantly increased levels of Arid3a in OE-High cells led to primarily nuclear localization (Figure 2.11A). Notably, differentiated ES cells, which have an elevated level of endogenous Arid3a, also showed enhanced nuclear accumulation of Arid3a (Figure 2.11B), as did TS cells (Figure 2.11C), which express >60-fold higher levels of Arid3a than ES cells (Figure 2.11D). These results indicated that Arid3a, when elevated to levels sufficient to initiate *trans*-differentiation of ES cells to TS-like cells (Figure 2.3C,2.6A), accumulates within the nucleus through either enhanced nuclear entry or reduced nuclear export.

Even though Arid3a resides primarily in the cytoplasm in normal or Arid3a OE-Low cells, we observed no evidence for Oct4 sequestration within the cytoplasm (Figure 2.11A). Thus, partitioning of Arid3a and Oct4 into different subcellular compartments in ES cells may explain why Arid3a knockdown had no effect on ES cell self-renewal

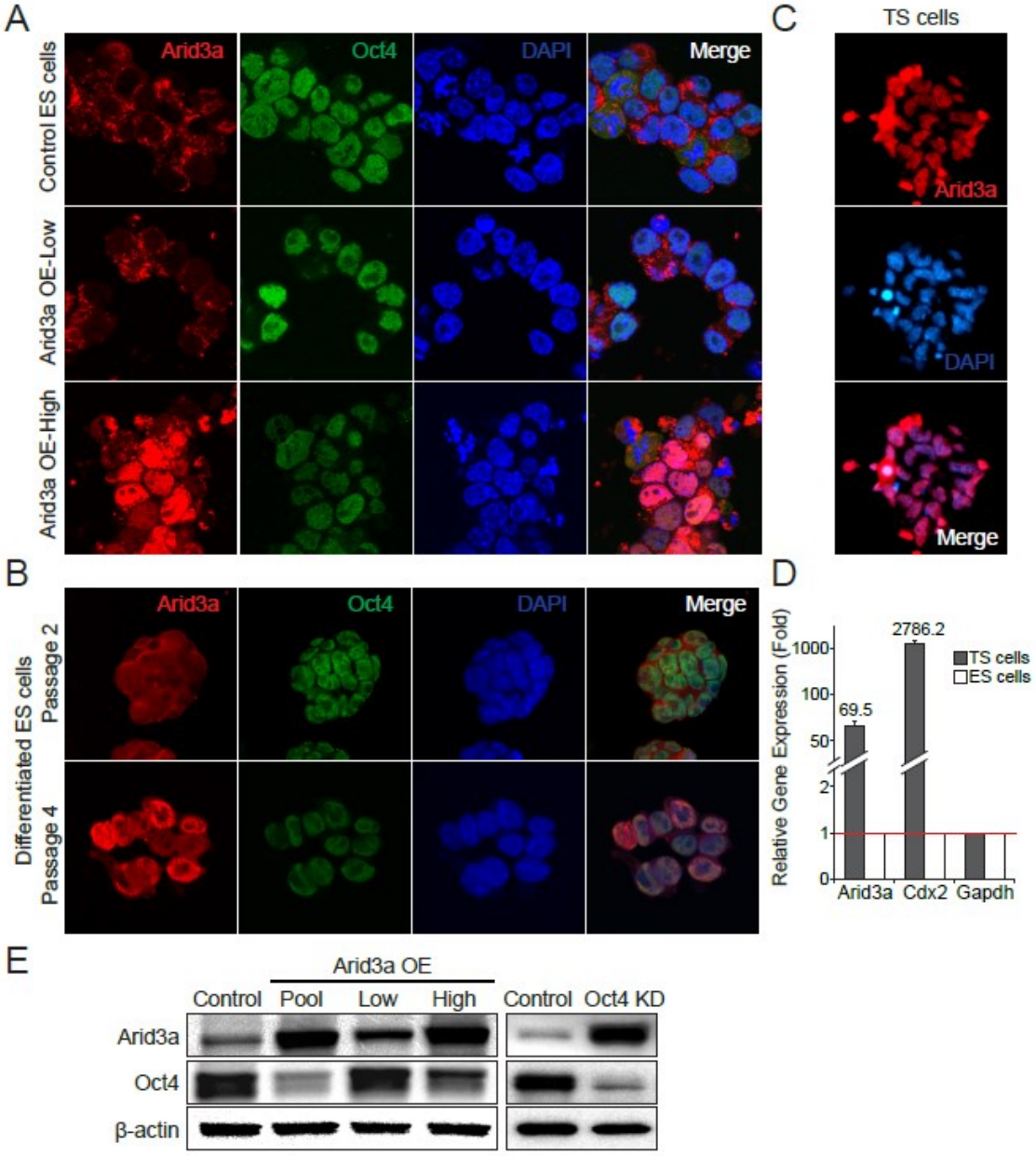


Figure 2.11. Nuclear localization of Arid3a is required for Oct4 repression and TE differentiation. (A) Immunofluorescence (IF) images of control ES, Arid3a OE-Low, and Arid3a OE-High cells. All cells were stained with directly conjugated anti-Arid3a (red, Alexa 594) and anti-Oct4 (green, Alexa 488) as well as DAPI (blue). Arid3a is primarily cytoplasmic in ES or Arid3a OE-Low cells but shuttles to the nuclei of Arid3a OE-High ES cells, resulting in reduced levels of nuclear Oct4. (B) IF images of ES cells upon differentiation (LIF-) conditions described in Materials and Methods. Each was stained with anti-Arid3a (Alexa 594), anti-Oct4 (Alexa 488), and DAPI (blue). (C) IF images of TS cells stained for endogenous Arid3a (red, Alexa 594) and DAPI (blue). (D) Relative expression levels of Arid3a and Cdx2 in bona fide TS cells and ES cells as determined by RT-qPCR. (E) Western blotting showing protein levels of Arid3a and Oct4 following Arid3a OE or Oct4 KD in ES cells. β-actin is the loading control.

(Figure 2.1A,B). Instead, immunofluorescence detected an opposite shift in nuclear intensities of Arid3a and Oct4 on OE of Arid3a (Figure 2.11A,B). This was confirmed semi-quantitatively in Western blots of whole-cell extracts; i.e., Oct4 was reduced ~50% relative to control ES cells (Figure 2.11E). That Oct4 knockdown up-regulated Arid3a levels by approximately eightfold (Figure 2.11E) further suggested that Arid3a and Oct4 may reciprocally repress one another.

Cytoplasmic-to-nuclear translocation of Tead4, in addition to gene activation of key members of the Hippo pathway, is required for TE specification (Nishioka et al., 2008). While mRNA and protein levels of Hippo pathway components were not significantly altered (Figure 2.12A,B), we observed enhanced nuclear accumulation of Tead4 in Arid3a OE-high cells relative to ES cells (Figure 2.12C). Thus, induction of nuclear Arid3a leads to differential Tead4 localization, potentially contributing to ES-to-TS cell *trans*-differentiation. Unlike Tead4, Yap1 remained primarily concentrated in the nucleus of ES, TS, and Arid3a-overexpressing cells (Figure 2.12C).

Target occupancy of Arid3a correlates with Arid3a-mediated gene expression

Arid3a and Oct4 reciprocally regulate transcript levels of many lineage-specific and pluripotency-associated genes (Figure 2.5A). This along with the above results (Figure 2.11) raised the possibility that Arid3a and Oct4 may share common target genes. We first mapped the global target loci of Arid3a using *in vivo* biotinylation-mediated chromatin immunoprecipitation (bioChIP) (Kim et al., 2009) followed by massively parallel sequencing. Higher levels of Arid3a OE, as expected from its enhanced nuclear localization (Figure 2.11), increased the overall occupancy signals of Arid3a (Figure 2.13A). The target loci of Arid3a were primarily distributed within intergenic (~42%) and promoters (~31%), suggesting that Arid3a regulates both proximal and distal regulatory elements (Figure 2.13B).We then compared the target occupancy of Arid3a with the gene expression profiles of Arid3a-overexpressing cells using a moving window average (window size, 250; bin size, 1). As shown in Figure 2.13C, a large number of genes up-regulated or down-regulated upon OE of Arid3a are direct transcriptional targets of

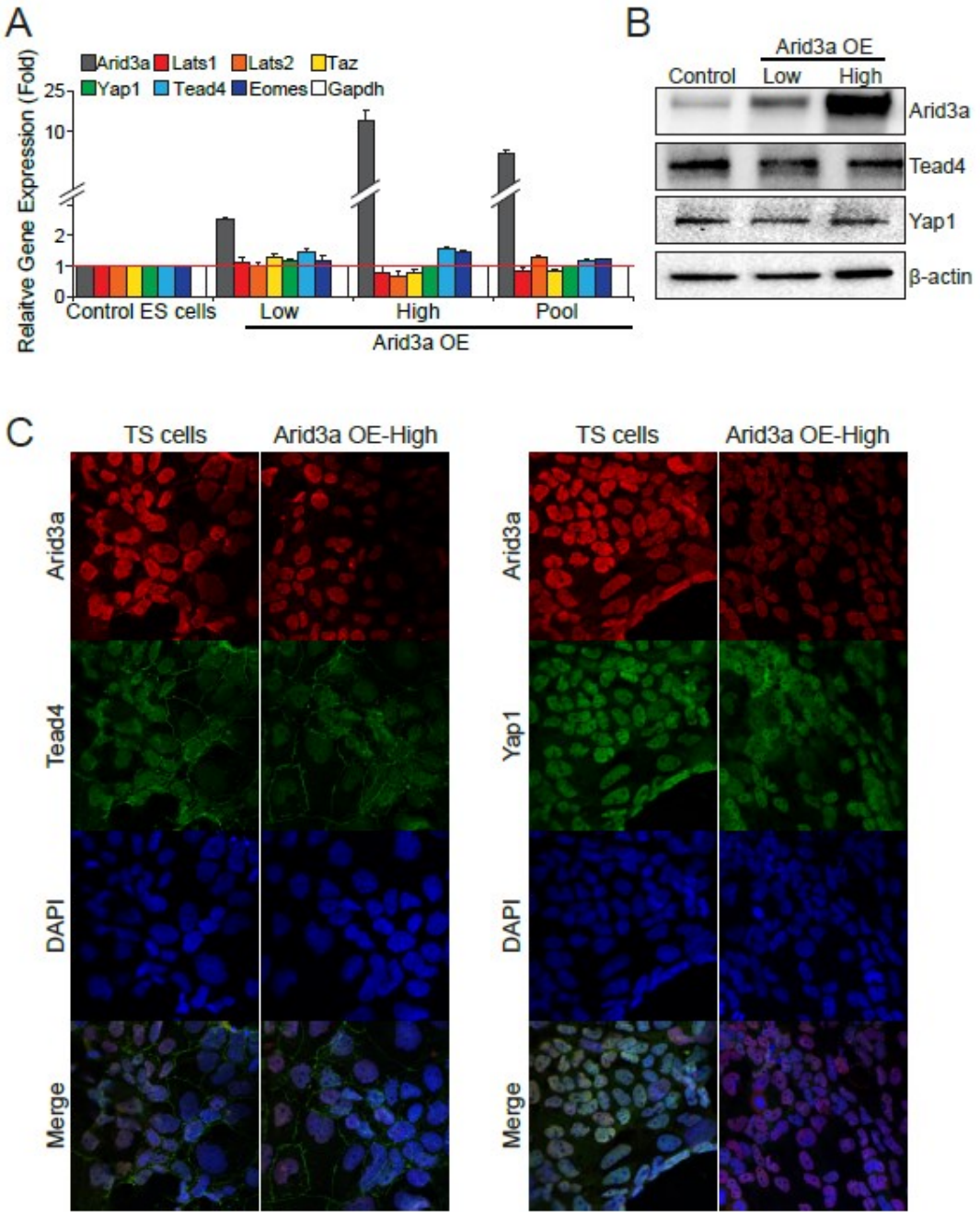


Figure 2.12. Induction of nuclear Arid3a leads to differential Tead4 localization. (A) Expression analysis of key Hippo pathway transcripts by RT-qPCR using control ES cells and Arid3a OE clones (Arid3a OE-Low, -High, and –Pool). Error bars depict standard deviations of biological triplicates. (B) Protein levels of key regulators of the Hippo pathway (Tead4 and Yap1), following Arid3a OE in ES cells as measured by Western blotting. β-actin served as a loading control. (C) Nuclear localization of Tead4 (left) but not Yap1 (right) correlates with high Arid3a nuclear levels in TS and Arid3a OE-High cells. For IF images, cells were stained with anti-Arid3a (red, Alexa594), anti-Tead4 or anti-Yap1 (green, Alexa 488), and DAPI (blue).

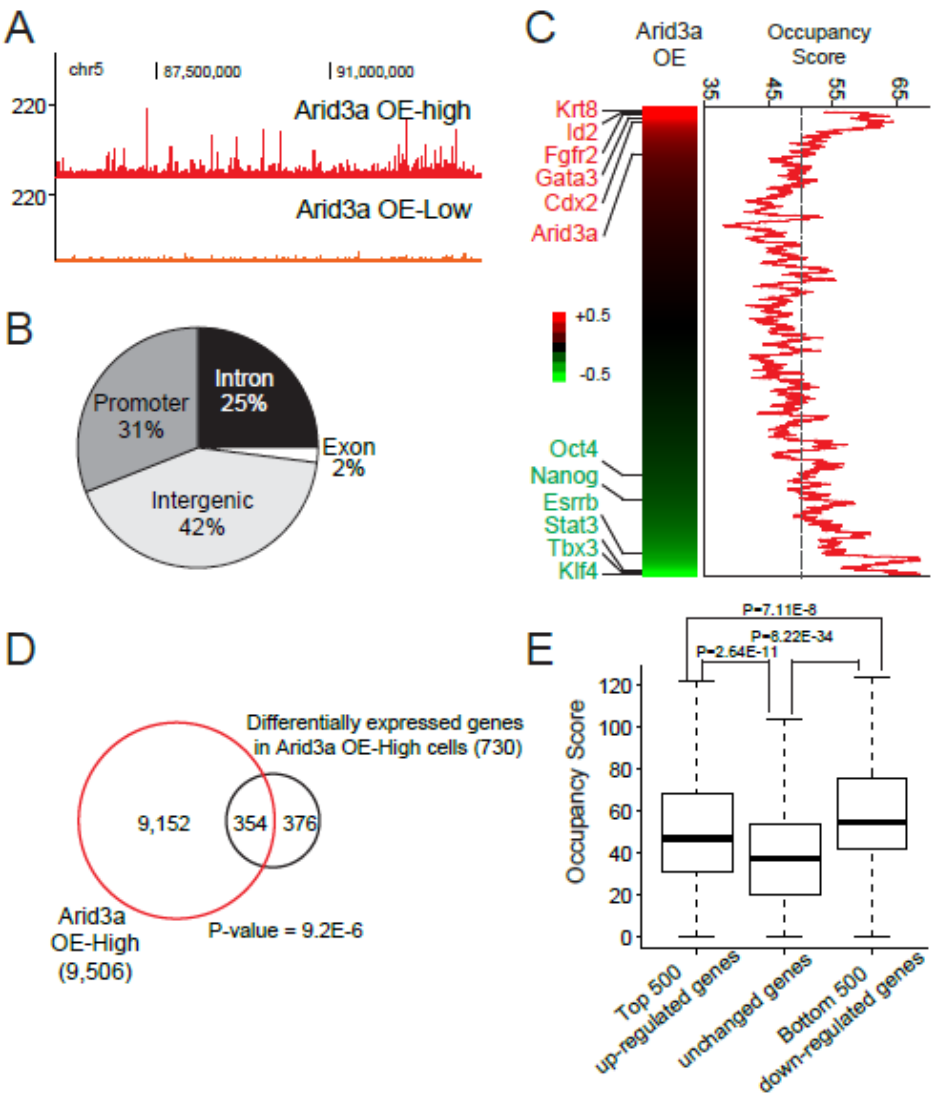


Figure 2.13. Target occupancy of Arid3a correlates with Arid3a-mediated gene expression. (A) A Snapshot of a random locus on chromosome 5 illustrating that Arid3a ChIP-seq occupancy signals derived from Arid3a OE-High ES cells are stronger than those of Arid3a OE-Low cells. (B) Pie chart presenting the distribution of Arid3a peaks. Promoters indicate regions within ±2Kb from the transcription start sites (TSS); Intergenic depicts regions other than promoters, exons, or introns. (C) A heatmap representation of the expression profile of Arid3a OE cells with representative gene names (red: TE-specific genes, green: pluripotency-associated genes). Genes were ordered according to gene expression levels in Arid3a OE cells relative to control ES cells. Moving average (window size: 250, bin size: 1) was plotted to corresponding Arid3a occupancy signals. (D) A Venn diagram showing overlap between Arid3a chromosomal target genes (red) and differentially expressed transcripts in Arid3a OE-High ES cells (black, 1.5-fold changes compared to control ES cells). (E) A box plot showing the distribution of Arid3a occupancy upon OE of Arid3a in ES cells of (left) the top 500 up-regulated genes, (middle) 500 genes without significant expression changes, and (right) the top 500 down-regulated genes.

Arid3a (Figure 2.13C-E). Thus, during Arid3a-mediated TE *trans*-differentiation, Arid3a acts as both an activator and a repressor of target gene transcription.

Arid3a directly activates TE markers and directly represses pluripotency factors during ES-to-TS cell *trans*-differentiation

Recent target mapping of various TFs in ES cells revealed that clusters of relatively high numbers of TFs act together to regulate their common targets (Kim et al., 2008). To gain further insight into Arid3a-mediated global gene regulation, we compared its target occupancy in Arid3a OE-High ES cells with publicly available ES cell ChIP-seq (ChIP combined with deep sequencing) data for the following factors: pluripotency-associated TFs (Oct4, Sox2, and Nanog), key components of Polycomb-repressive complex (Ezh2, Suz12, and Ring1b), Myc-related (n-Myc and E2f1), and the TE master regulator Cdx2 (Figure 2.14A). Although Arid3a and Cdx2 have similar consequences on ES-to-TS cell *trans*-differentiation, we found very low accordance in their target sets (Figure 2.14A-C), indicating that their underlying TE regulatory mechanisms must differ. Instead, Arid3a target loci strongly overlapped with those of pluripotency-associated factors (Figure 2.14A-D), and Arid3a target genes are highly enriched in the previously defined ES cell core module (Figure 2.14E)(Kim et al., 2010). Notably, we observed strong enrichment of the Oct4 DNA-binding motif in Arid3a target loci with the perfectly conserved core octamer sequence (ATGCAAAT) (Figure 2.14F). Accordingly, Arid3a and Oct4 occupied the same regulatory regions of numerous pluripotency-associated genes, including *Oct4* itself and *Nanog*, whereas TE-specific genes, including *Cdx2, Gata3, Id2*, and *Tcfap2c*, were singularly occupied by Arid3a (Figure 2.14G,H). Gene ontogeny analyses predicted highly significant enrichment of Arid3a targets in signatures associated with blastocyst formation, trophectodermal cell differentiation, and early embryogenesis (Figure 2.14I,J). We conclude that, in general, the gene targets shared by Arid3a and Oct4 are repressed, whereas the targets occupied by only Arid3a are activated upon induction of Arid3a in ES cells.

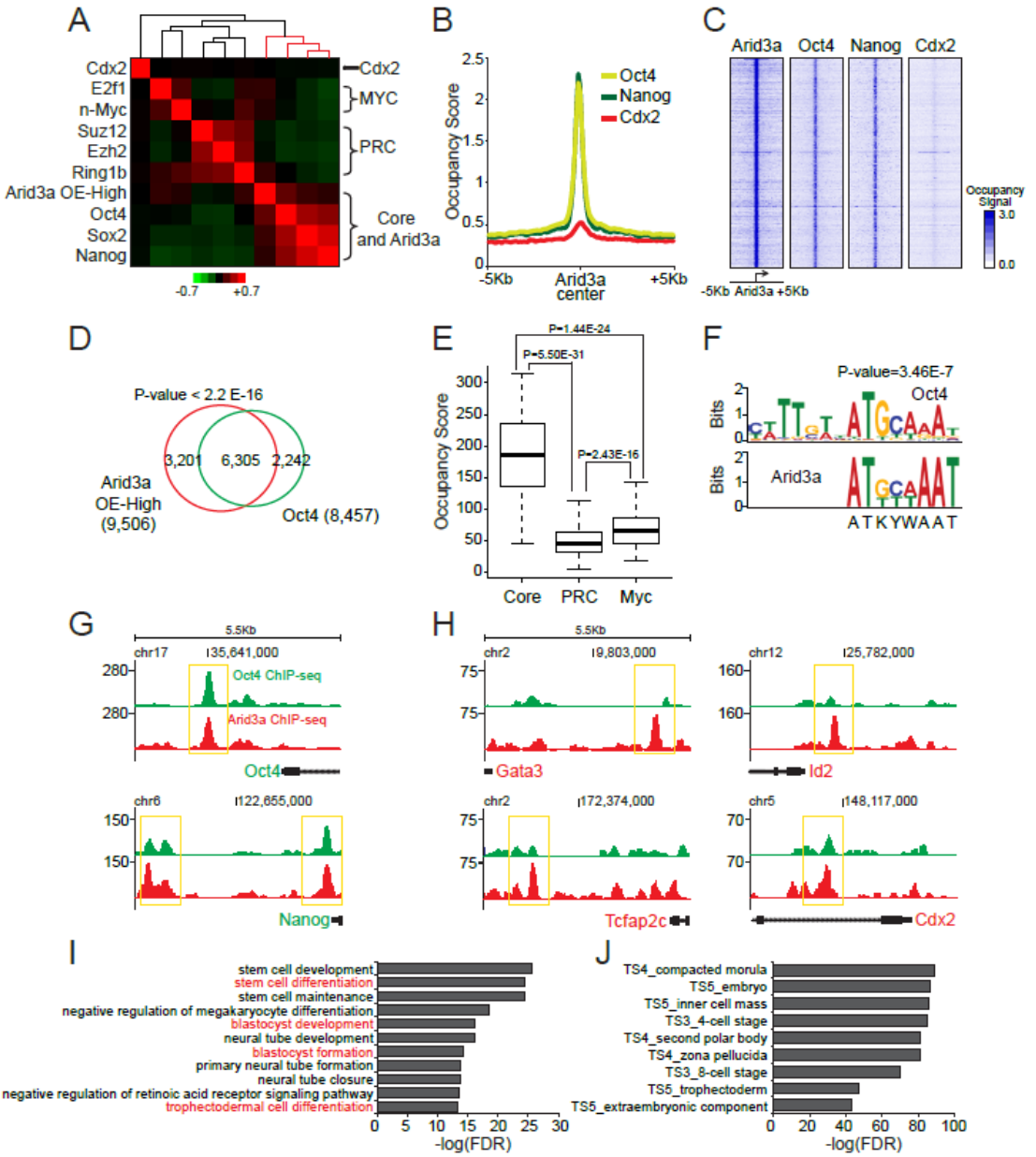


Figure 2.14. Arid3a and Oct4 occupy highly similar target loci. (A) Target correlation map of indicated TFs illustrating that Arid3a target loci overlap with those of core TFs. (B) Binding positions of Oct4 (yellow), Nanog (green), and Cdx2 (red) relative to the center of corresponding Arid3a target loci. (C) Read-based ChIP-seq signals of Oct4, Nanog, and Cdx2 at the Arid3a target sites (± 5Kb). (D) Venn diagram showing that the majority of Oct4 and Arid3a target genes overlap. (E) A box plot showing the distribution of Arid3a occupancy signals for the previously defined three pluripotency modules (Core, PRC, and Myc) (Kim et al. 2010). (F) Motif analysis of Arid3a and Oct4 target loci indicates that the prototypical octamer motif is the most highly enriched DNA binding site for both TFs. (G,H) Snapshots of ChIP-seq signal tracts of Oct4 (green) and Arid3a (red) at the regulatory regions of pluripotency-associated genes (G) and TE markers (H). (I,J) Enriched biological functions (I) and MGI expression terms (J) of the top 5,000 strongest Arid3a target peaks by Gene Ontology (GO) analyses.

Arid3a represses pluripotency-associated genes by recruiting HDACs

HDACs have been implicated in numerous transcriptional repression and chromatin remodeling contexts as well as in silencing of cell fate decisions (Dovey et al., 2010). Recent reports suggested that HDACs are associated with ES cell core factors in a complex termed ‘‘NODE’’ for Nanog and Oct4-associated deacetylases (Liang et al., 2008). Furthermore, a significant number of pluripotency-associated TFs share their targets with HDACs (Whyte et al., 2012). Thus, we reasoned that HDACs may distinguish Arid3a-mediated repression from Arid3a-mediated activation (Figure 2.14G,H). Consistent with this hypothesis, we observed strong Arid3a–HDAC1/2 co-immunoprecipitation (co-IP) in Arid3a OE-High cells (Figure 2.15A).

Since the levels of HDAC1 and HDAC2 were not elevated by Arid3a OE in ES cells (data not shown), we used HDAC1 ChIP to determine whether the target occupancies of HDAC1 are altered upon OE of Arid3a. As shown in Figure 2.15B, HDAC1 occupied regulatory elements of pluripotency-associated genes more strongly following Arid3a OE than in control ES cells, whereas only modest changes were observed in regulatory elements of TE-specific genes. In contrast, activation-associated histone acetylation marks (AcH3 and AcH4) were significantly enriched at the regulatory regions of TE-specific marker genes, whereas weaker signals were associated with pluripotency-associated genes (Figure 2.15C). These data suggested that the repression of pluripotency-related genes in Arid3a-overexpressing cells is at least in part mediated by the activity of HDAC1 and Arid3a, whereas histone acetyltransferases (HATs) are recruited to TE-specific regulatory elements with Arid3a. In further support of this notion, target genes co-occupied by both HDAC1 and Arid3a, many of which are ES cell-specific (Figure 2.15F, top), are strongly repressed upon induction of Arid3a (Figure 2.15D,F). Conversely, regulatory elements of TE-specific genes that are up-regulated upon OE of Arid3a were not co-occupied by HDAC1 and Arid3a (Figure 2.15E). Unique targets of HDAC1 were not restricted to either TS cell- or ES cell-specific genes (Figure 2.15F, bottom). These data along with the co-IP results (Figure 2.15A) indicate that the

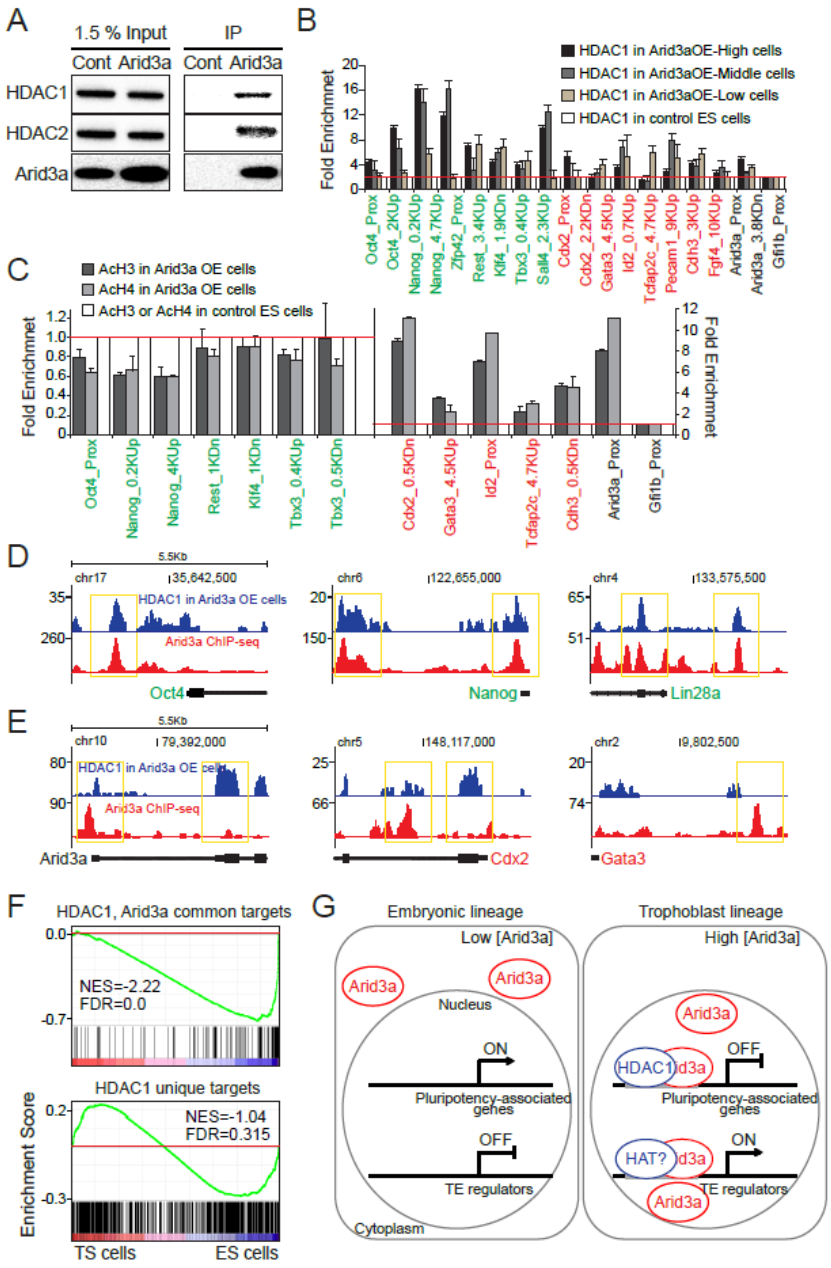


Figure 2.15. Arid3a recruits HDACs onto loci of pluripotency-associated genes. (A) Immunoprecipitation (IP) of Arid3a complexes from control ES cells (Cont) or Arid3a OE-High ES cells (Arid3a). (B) Fold enrichments of HDAC1 binding to the regulatory regions of pluripotency-associated genes (red), lineage markers (green), and Arid3a (black). (C) Fold enrichments of AcH3 (gray) and AcH4 (light gray) marks at the regulatory regions of pluripotency-associated genes (red), lineage markers (green), and Arid3a (black). Error bars depict standard deviations of biological triplicates (B,C). Prox, Up, and Dn indicate regions proximal, upstream, and downstream, respectively, of the TSS (B,C). (D,E) Snapshots of ChIP-seq signal tracts of HDAC1 (blue) and Arid3a (red). (F) GSEA using gene sets of targets in common to HDAC1 and Arid3a or unique to HDAC1. (G) Embryonic vs trophoblastic lineage choice depends on differential nuclear import and HDAC association of Arid3a.

repression of pluripotency-associated genes upon OE of Arid3a is mediated by selective recruitment of HDAC1 onto their common *cis*-regulatory elements.

2.4 DISCUSSION

Here we identify Arid3a as a central transcriptional regulator of TE lineage commitment and differentiation. Arid3a is highly expressed *in vitro* in nuclei of TE and TS cells as well *in vivo* in outer cells of early blastocyst and TE-derived placental lineages. OE of Arid3a is sufficient to transdifferentiate ES cells into TS-like cells *in vitro*. When transplanted into developing mouse blastocysts *ex vivo*, Arid3a-driven TS-like cells integrate into the TE layer of developing embryos. The induction of Arid3a in TS cells leads to further differentiation of TS cells to more specialized trophoblastic lineages—a fate not executed by Cdx2. Arid3a acts both directly upstream and independently of Cdx2 to promote TE pathways while directly repressing Oct4-dependent pluripotency pathways. We provide evidence that selective TE versus ES gene regulation is mediated epigenetically via differential HDAC recruitment.

Trans-differentiation of ES cells toward TE lineage via OE of a single TF was previously reported for Cdx2 and Gata3 (Niwa et al., 2005; Ralston et al., 2010). Although our data on Arid3a show similar consequences on TE fate specification, we found that Arid3a and Cdx2 do not appear to share similar regulatory mechanisms. While the prior work using Cdx2-inducible ES cells showed exclusive target occupancy of Cdx2 on TE lineage-specific genes (Nishiyama et al., 2009), our genome-wide mapping revealed that Arid3a directly binds to the regulatory elements of both TE-specific factors (including *Cdx2, Gata3, Tcfap2c*, and *Id2*) and pluripotency-associated genes (such as *Oct4* and *Nanog*) during *trans*-differentiation (Figure 2.13). This differential target gene occupancy between Cdx2 and Arid3a is consistent with their phenotypic dissimilarity found during the differentiation of TS cells into more specialized trophoblast cells (Figure 2.6), further indicating the existence of multiple sub-regulatory networks responsible for TE lineage specification or differentiation.

Notably, we found that Arid3a directly occupies the regulatory elements of Cdx2 (Figure 2.14), suggesting that, at least in certain contexts, Arid3a is an upstream regulator of this TE ‘‘master’’ regulator. On the other hand, Cdx2 and Arid3a also operate in parallel pathways to achieve TE *trans*-differentiation, as shown by reciprocal knockdown and OE analyses of Figure 2.7. It will be of great interest to identify additional key regulators of TE lineage specification and elucidate how these regulators are intertwined in combination with other co-regulators to coordinate downstream gene expression programs during early cell fate decision and eventually placenta development.

We show TE segregation both *in vivo* and in cultured ES cells under natural or ectopically induced high levels of Arid3a, which the subcellular localization of Arid3a correlates strongly with execution of TE fate. Arid3a largely resides in the cytoplasm of early blastomeres and the ICM but, by E4.5, undergoes translocation to nuclei of the outer TE cells (Figure 2.10,2.11). Similarly, its translocation from self-renewing ES cell cytoplasm to the nucleus is required to repress the TE antagonist Oct4 to levels sufficient to promote ES-to-TE *trans*-differentiation and consistent with prior studies (Niwa et al., 2000). Further supporting our hypothesis, both TS cells and Arid3a OE-derived TS-like cells as well as placental TGCs and SpTs displayed primarily nuclear localization of Arid3a (Figure 2.9,2.10). Previous studies showed that Arid3a undergoes CRM1-dependent nucleocytoplasmic shuttling in somatic cell types, but the underlying mechanism remains unknown (Kim and Tucker, 2006). A similar mechanism was shown to activate Tead4, the well-defined Hippo pathway TF essential for TE specification (Nishioka et al., 2008). Accordingly, we observed Tead4 translocation correlated with Arid3a nuclear levels (Figure 2.12). An important subject of future study will be to determine how the nuclear localization of Arid3a is post-translationally regulated and the consequences on its distinct transcriptional programs in ES and TS cells.

We observed a strong correlation between Arid3abound target genes with transcripts deregulated upon Arid3a OE. The general pattern was that upon Arid3a OE, TE factors were activated, and pluripotent factors were repressed. A central question raised by our

integrative DNA-binding and global gene expression analysis was how Arid3a functions as both an activator and a repressor depending on its target genes. The first hint was our finding that Arid3a interacts with HDAC1 and HDAC2 at the protein level (Figure 2.15). HDACs generally convey repressive function in somatic cells, particularly in the context of the NuRD complex. However, recent studies indicate that HDACs are also bound to active chromatin in both ES and TS cells (Kidder and Palmer, 2012). In ES cells, HDAC1 associates with regulatory regions of key TFs implicated in the maintenance of ES cell self-renewal (such as *Oct4, Sox2,* and *Nanog*), whereas, in TS cells, HDAC1 binds to crucial TE genes (including *Cdx2, Elf5*, and *Eomes*) (Kidder and Palmer, 2012). Our HDAC1 ChIP followed by quantitative PCR (qPCR) analysis indicated that Arid3a OE promotes stronger occupancy of HDAC1 onto the regulatory elements of pluripotency-associated genes without significant occupancy changes at the regulatory elements of TE-specific genes. Importantly, a high level of Arid3a–HDAC1 co-occupancy is observed at pluripotency-associated loci in Arid3a-overexpressing TS-like cells. Perhaps this Arid3a–HDAC1 interaction alters the activity/specificity of HDAC1, leading to deacetylation of pluripotency-associated nucleosomal histone tails (Figure 2.15). Thus, differential co-occupancy of HDAC1 and Arid3a may be responsible for the repressive role of Arid3a. Our results in Figure 14D additionally suggest putative associations of Arid3a and HAT for the activation of TE-specific genes.

While Arid3a can phenocopy, if not surpass, Cdx2 in executing ES-to-TE *trans*-differentiation in cultured cells, there is a critical difference *in vivo*. Cdx2 is critical for the first cell fate commitment of the TE, as *Cdx2*$^{-/-}$ embryos fail to implant (Strumpf et al., 2005). However, while fetal death was observed as early as E8.5 (Figure 2.9A), the majority of *Arid3a*-null embryos can implant and develop to as late as E11.5 (Webb et al., 2011), albeit with gross defects in placental cellularity and architecture, including expression and organization of the direct descendants of TE, TGCs, and SpTs (Figure 2.9)(Uy et al., 2002). However, in the *in vitro* context, Arid3a can directly activate *Cdx2* and indirectly activate an arm of the Hippo pathway (shown to be upstream of Cdx2 by inducing Tead4 nuclear localization)—events associated previously with the specification

phase of the first cell fate. Perhaps, *in vitro*, the distinction between specification and commitment is less strict than *in vivo*, where these steps are readily distinguished by preimplantation and post-implantation. We are currently examining *Arid3a*$^{-/-}$ preimplantation embryos to better resolve this issue.

Although not entirely analogous, many of these processes in mice are conserved in human placenta development. A notable exception and in support of the possibility that Arid3a acts in parallel with Cdx2 is that the reciprocal Cdx2–Oct4 expression patterns observed in mice are not conserved in humans (Hay et al., 2004). Further consistent with defects in the *Arid3a*$^{-/-}$ placentas, disruption of the migration of differentiating TGCs can result in poorly formed vessels that lead to vascular insufficiency of the placenta and fetus, which in turn may result in preeclampsia (Red-Horse et al., 2004). Expression profiling of human chorionic trophoblast progenitor cells and primary villous trophoblasts identified Arid3a as highly expressed in these lineages (Genbacev et al., 2011). Human TGCs, which highly express Arid3a (Genbacev et al., 2011), are the first cell type to terminally differentiate during embryogenesis and are of vital importance for implantation and modulation of the post-implantation placenta (Red-Horse et al., 2004). These findings point to potential Arid3a involvement in human placental fate and maintenance and suggest that it will be extremely informative to determine Arid3a function in both normal and preeclamptic placenta.

CHAPTER 3: ARID3A IS REQUIRED FOR MAMMALIAN PLACENTA DEVELOPMENT

3.1 INTRODUCTION

Two sequential cell fate decisions during blastocyst formation establish three distinct cellular lineages—trophectoderm (TE), epiblast, and primitive endoderm (Rossant and Tam, 2009; Zernicka-Goetz, 2004). The first cell fate decision in mice and humans occurs at the 8- to 16-cell stage and leads to segregation of the inner cell mass (ICM), which gives rise to all tissues of the body, and the TE, which is required for implantation into the uterus and formation of the placenta (Niwa, 2007). The placenta is essential for survival of the mammalian embryo, as it transports nutrients, produces hormones, provides structural support within the womb, provides immunological protection, and acts as a physiological buffer between the mother and the fetus (Simister and Story, 1997). Abnormal placental development underlies a wide range of complications during pregnancy, including preeclampsia (PE), miscarriage, and predisposition to chronic disease in adulthood (Roberts et al., 1989; Suzuki, 2008). PE, a pregnancy-specific placental disorder characterized by the development of hypertension during gestation, is a major obstetric problem that contributes substantially to maternal and perinatal morbidity and mortality worldwide (Ananth et al., 2013).

While genome-wide analyses have identified genes deregulated in PE, only a few transcription factors (TFs) have been shown to be associated with normal TE specification and/or human placental differentiation (Hemberger et al., 2010; Martinez-Fierro et al., 2016). Its strong similarity with human placentation makes the mouse an excellent model to elucidate key mechanisms of placental development.

Our previous studies showed that AT-Rich Interactive Domain 3A (ARID3A) is

---

Reproduced in part from Catherine Rhee, Melissa Edwards, Christine Dang, June Harris, Mark Brown, Jonghwan Kim, and Haley O. Tucker (2016). ARID3A is required for mammalian placenta development. *Dev Biol.* In press. (Contributions: C.R. and H.O.T. designed and analyzed the experiments. C.R., M.E., and J.H. performed mouse experiments. C.R., and C.D. performed human iPS cell experiments. C.R. and H.O.T. wrote the manuscript with input from other authors)

essential for the first cell fate decision (Rhee et al., 2014). We found that overexpression (OE) of ARID3A alone is sufficient for trans-differentiation of embryonic stem (ES) cells to trophoblast stem (TS)-like cells—the *in vitro* counterpart of the TE layer of the blastocyst. Global expression profiles of ARID3A-OE ES cells and TS cells are highly similar. Arid3a-OE ES cells gain the capacity to incorporate into the TE of developing embryos—an indication of the development of functional TS cells.

To gain insight into the role of Arid3a in placentation, we have carried out further *in vivo* analyses in the mouse and *in vitro* analyses in the human. Our data indicate that ARID3A provides an indispensable and conserved function in mammalian placental development and may provide a novel diagnostic marker for PE.

3.2 MATERIALS AND METHODS

Mouse TS cell culture

Mouse TS cells were maintained at a ratio of 3:7 of TS medium to MEF-conditioned TS medium, supplemented with 25ng/mL Fgf4 and 1ug/mL of Heparin (TS+ medium). TS medium is RPMI 1640 (Roswell Park Memorial Institute Medium, Gibco) supplemented with 20% FBS, 100uM β-mercaptoethanol, 2mM L-glutamine, 1mM sodium pyruvate, penicillin (50U/mL), and streptomycin (50mg/ML). MEF-conditioned medium is TS medium conditioned by MEF. MEF were treated with mitomycin, followed by culture for 9 days with collecting the medium every third days.

Mouse TS cell *in vitro* differentiation

Mouse TS cells were differentiated for four days by withdrawing Fgf4 and Heparin from the TS+ medium described above.

Human iPS cell culture

Human iPS cells (Yamanaka retrovirus reprogrammed hiPSC, ACS-1023, ATCC) were passed for sterility tests, showing negative results for bacterial, fungal, and mycoplasma, and STR profile of the cells were confirmed by STR analysis and karyotype by G-banding (performed by ATCC). The cells were maintained in mTeSR1 media (Stemcell technologies). iPS cells were cultured in matrigel (corning) plates. The cells were

incubated at 37°C, 5% CO_2. For individualization of iPS cells, the cells were detached using TrypLE followed by plating with mTESR1 supplemented with Y27632 (ROCK inhibitor, Alexis biochemical).

Human TS-like cell generation

Human TS-like cells were generated by adding 250ng/mL of BMP4 to human iPS cell culture medium as described above for 48 hours.

Lentiviral production and infection

293T cells (Sigma) were used to produce lentivirus for shRNA-based KD of ARID3A with pCMV-Δ8.9 and VSVG helper plasmids using lipofectamine 3000 (Invitrogen), according to the manufacturer's instructions. After 12 hours, 293T medium was replaced with human iPS medium (mTeSR1). 48 hours following transfection, the supernatant-containing viral particles were filter-collected.

Immunohistochemistry

Tissue sections were deparaffinized, permeabilized with 100 ul of 0.1% Tween20, then treated with 3% H_2O_2 to block endogenous peroxidase activity. Antigens were retrieved by adding boiling 10 mM citrate buffer (pH 6) and incubating the sections in a 95°C water bath for 20 minutes. Slides were then washed with 0.01% Tween20, 1% BSA in 1X Dulbecco's PBS, and blocked with 10% horse serum, 1% BSA in TBS for 1 hour. Primary antibody was diluted in TBS + 1% BSA, and slides were incubated at 4°C overnight; secondary antibody was diluted in the same manner. Sections were then developed using Thermo Scientific Pierce DAB substrate (Fisher Scientific), counterstained with Haemotoxylin (Fisher Scientific), mounted with organo-limonene mounting media (Sigma-Aldrich), and stained with anti-Arid3a (Webb et al., 2011), anti-Hand1 (Santa Cruz, sc-9413), anti-Proliferin (Santa Cruz, sc-47347), anti-Pl1 (Santa Cruz, sc-376436), anti-Mash2 (GeneTex, GTX60272), anti-Cdx2 (Abcam, ab88129), anti-Gcm1 (Thermo Fisher, PA5-46853) antibodies.

RNA isolation and RT-qPCR

RNA was isolated using the RNeasy Mini Kit (Qiagen). 500ng RNA was then used to synthesize cDNA with qScript cDNA supermix (Quanta). 2uL of 20X diluted cDNA was

used per well of a 96-well plate for RT-qPCR analyses performed using SYBR PerfeCta SYBR Green fast mix (Quanta). Primers for qPCR were designed to amplify a ~100 base pair region containing an overlap of two exons. The Ct values obtained from qPCR were normalized against Gapdh expression to determine relative gene expression.

Immunofluorescence (IF)

Cells were plated at high density on 6-well plates (Ibidi) and grown for 1-3 days at 37°C. Slides were fixed using 4% paraformaldehyde, followed by incubation in 0.5% Triton X-100 in blocking solution (3% BSA + 1% Normal Horse Serum in PBS). Slides then were incubated in Arid3a primary antibody solution (1:200 dilution) overnight at 4°C, and secondary Rabbit antibody conjugated to Alexa Flour 594 dyes (1:1000 dilution) for 1 hour in the dark at room temperature. Lastly, the plates were imaged on a Zeiss 710 Laser Scanning Confocal & SIM microscope. Anti-Arid3a affinity purified antibody (Webb et al., 2011) was used for staining.

Weighing and measuring of embryos

Embryos of Arid3a-/- null C57BL/6 mice (Webb et al., 2011) were obtained from timed matings of heterozygous Arid3a breedings and harvested at E10.5-E12.5. The uterine horns from the pregnant females were removed and each individual embryo was separated from the uterus with fine forceps and scissors before being placed in RPMI media. Under a dissecting microscope the yolk sack of each fetus was removed with fine forceps while leaving the placenta attached. The fetus and placenta were then flushed with 1x PBS prior to being weighed together. Each placenta was then detached from its fetus and individually weighed and measured. Weights were determined by subtracting the combined fetal and placental weights from those of the placenta. Measurements were determined by placing each placenta in a flat dish and taking the average of two maximum diameters at 90° angles with a metallic scale graduated in millimeters (mm). Numbers of embryos and placentas shown in Figure 2A were used for the measurements and statistical analysis shown in Figure 2C and D. All mice were euthanized under the guidelines of IACUC protocol (AUP-2015-00195) registered at our institution.

Semi-quantitative trophoblast giant cell (TGC) assessment

Since TGC regional disruption was severe in *Arid3a* nulls, TGC counting was assessed semi-quantitatively by visual inspection in the multiple section images.

RNA-sequencing analysis

Placentas of E10.5 and E11.5 *Arid3a*$^{+/+}$ and *Arid3a-/-* mice were analyzed by deep sequencing, using Illumina HiSeq 2500, paired-end technology. Placentas were removed, flash frozen on dry ice, and RNA was harvested using Trizol reagent. Total RNA was employed for library preparation using standard Illumina protocols. Once data were generated, Illumina software was used for base calling. Sequenced reads were trimmed for adaptor sequence, then mapped to mm9 whole genome using tophat. RPKM was calculated using 'cuffdiff'. Sequencing data are submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE79638.

3.3 RESULTS

Arid3a is highly expressed during mouse and human placentation

We first analyzed published global expression profiling for each stage of mouse embryonic development (Smith et al., 2014). *Arid3a* expression initiates in the morula, then later, becomes highly expressed in extraembryonic components (Figure 3.1A). Because a subset of these extraembryonic components are expressed in the placenta, we compared the levels of *Arid3a* with other key placental markers, including *Gata3*, *Tfap2c*, *Hand1*, and *Id2 in* vivo and upon time-course differentiation of TS cells-the *in vitro* counterpart of the TE (Kidder and Palmer, 2010). As shown in Figure 3.1A, both *Arid3a* and these transcripts are highly enriched within extraembryonic components and TE. However, other TE markers such as *Cdx2*, *Eomes*, *Id2*, and *Hand1*, are down-regulated upon TS differentiation (Figure 3.1B). This suggested that Arid3a differs from previously studied TE markers in that it expresses in mouse placentation both *in vivo* and *in vitro*.

Similar to the mouse, *ARID3A* levels in human are the highest in the placenta as compared to all other tissues (Li et al., 2013; Rhee et al., 2014)(Figure 3.2A). Analyses of

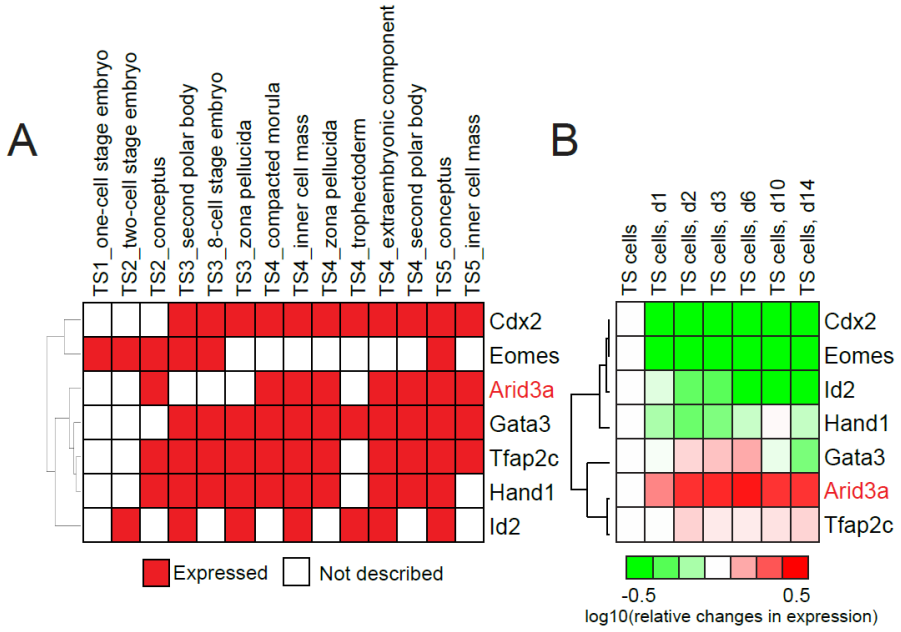


Figure 3.1. Arid3a is highly expressed during mouse placentation. (A) Global expression profiling of TE markers in early mouse embryonic developmental stages. (B) Expression levels of TE markers upon differentiation of mouse TS cells. Differentiation day, d

publically available data sets (Mikheev et al., 2008) further revealed that *ARID3A* is highly expressed throughout placental gestation from the 1st to 3rd trimester (Figure 3.2B) as well as in human TS-like cells induced from ES cells by BMP4 (Xu et al., 2002)(Figure 3.2C). Unlike BMP4-induced TS-like cells, neither human ES cell-derived endodermal cells (Figure 3.2D) nor embryoid bodies (Figure 3.2E) displayed significant induction of human *ARID3A*.

Loss of ARID3A during early mouse gestation results in intrauterine growth restriction and defects in placental development

Arid3a$^{-/-}$ null C57BL/6 mice (Webb et al., 2011) were employed to investigate the role of Arid3a in early placental development. Breeding heterozygous *Arid3a* mice resulted in non-Mendelian ratios from E10.5-E12.5, with no homozygous mutants obtained at E12.5 (Figure 3.3A). Since our previous studies (Rhee et al., 2014) detected high ARID3A expression in the TE, we carefully examined the gross anatomy of all E10.5 and E11.5 surviving embryos and placentas from *Arid3a* mutant heterozygous crosses. *Arid3a* mutants exhibited a range of phenotypes, from indistinguishable, to small, to paler

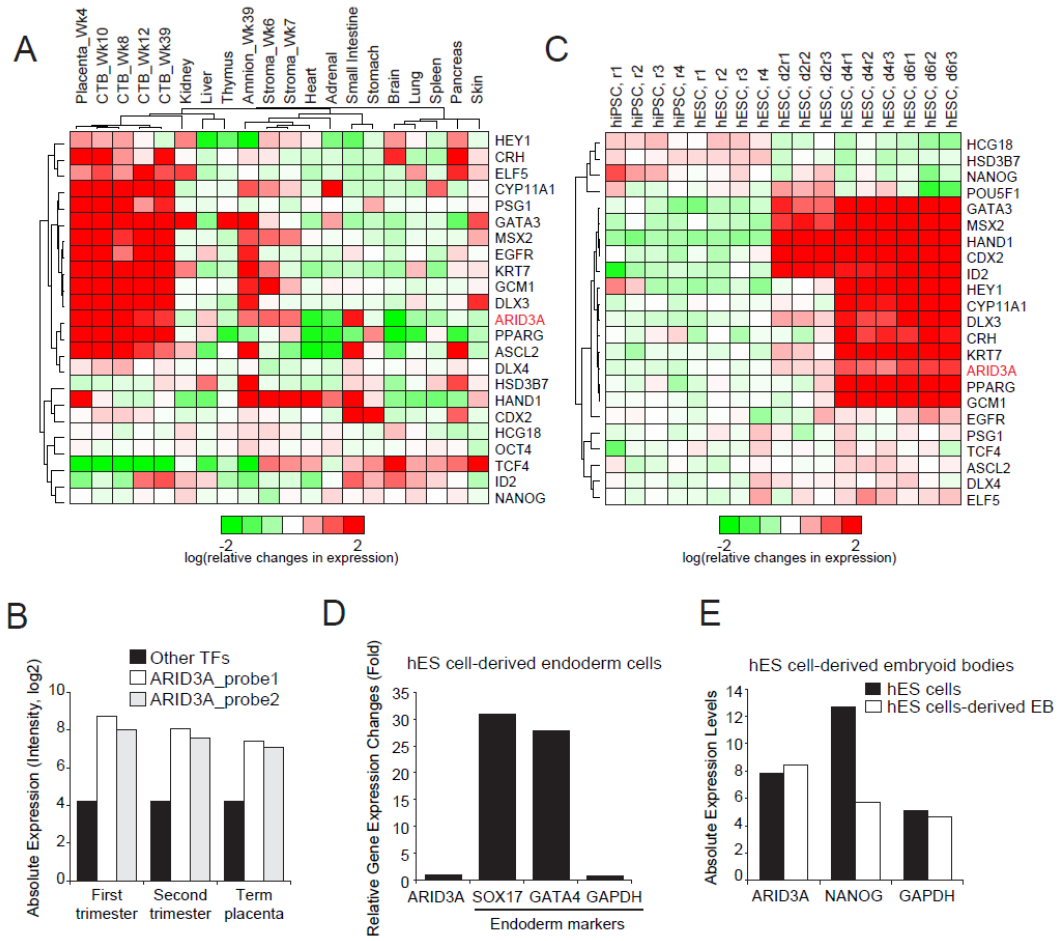
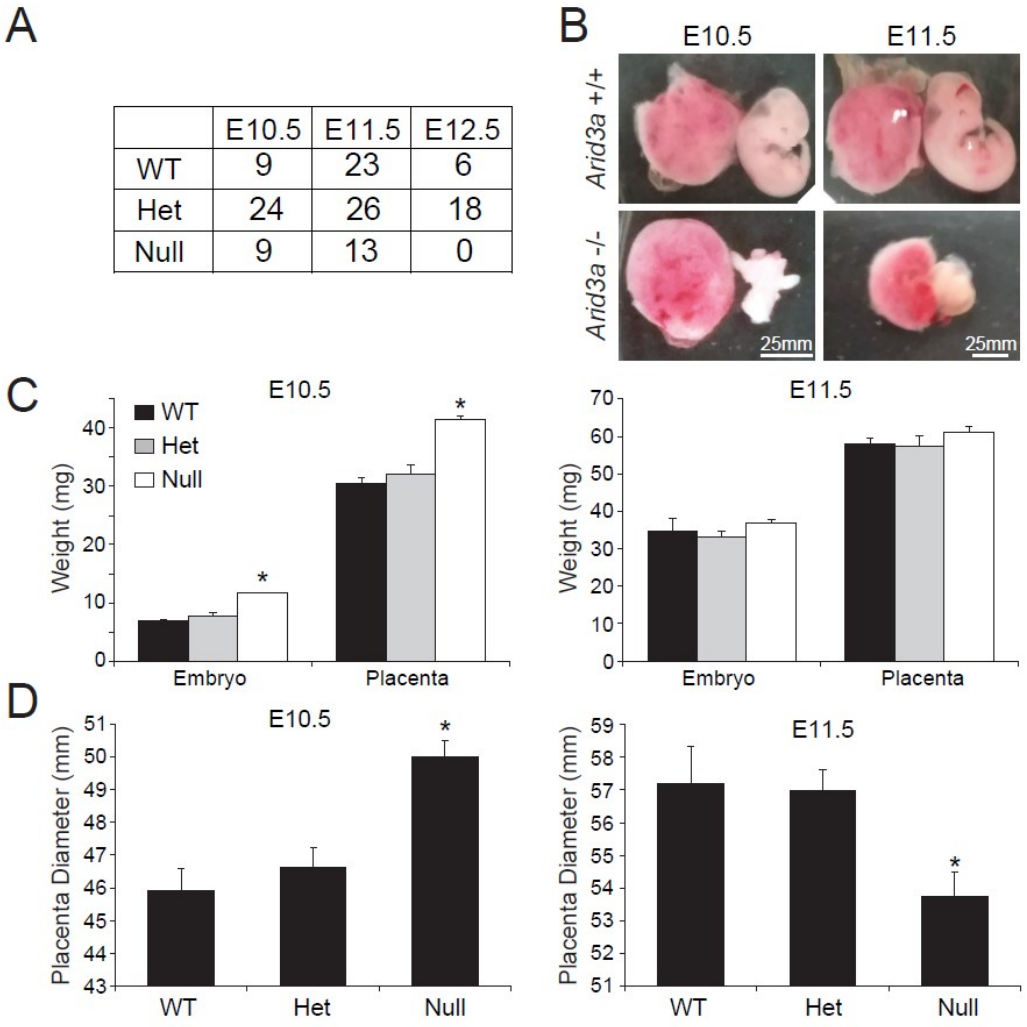


Figure 3.2. Arid3a is highly expressed during human placentation. (A) Global expression profiling of TE and pluripotency markers in human tissues, including placenta. (B) Absolute expression levels of ARID3A in placenta throughout gestation from 1st to 3rd trimester. Two ARID3A values were two different human ARID3A probes used in human microarray data. Two values were detected from two different regions of sequences within ARID3A coding DNA sequences. (C) Expression levels of TE and pluripotency markers upon BMP4-mediated differentiation of human ES and iPS to TS cells. Replication, r. (D) Expression levels of ARID3A in human ES cell-derived endoderm cells showing that ARID3A level is not altered. (E) Expression levels of ARID3A in human embryoid bodies showing that ARID3A level is not significantly altered upon differentiation.

embryos and placentas—further indications of the vasoconstriction characterized previously (Webb et al., 2011)(Figure 3.3B,C).

E10.5 nulls were frequently observed undergoing absorption (data not shown), and perhaps the adhesive properties underlying this as yet to be determined phenomenon account for their heavier weights relative to E10.5 wild-type (WT) controls (Figure 3.3C). Mutant placental diameters at E10.5 also significantly exceeded those of WTs but



	E10.5	E11.5	E12.5
WT	9	23	6
Het	24	26	18
Null	9	13	0

Figure 3.3. *Arid3a* KO in mouse embryos results in defects in placental development. (A) The numbers of viable placentas and embryos at E10.5, E11.5 and E12.5. Homozygous lethality of *Arid3a* null embryos was observed at E12.5. (B) Representative examples of *Arid3a* KO and WT mouse embryos and placentas at E10.5 and E11.5. (C) Weight comparisons among WT, heterozygous and null embryos and placentas. (D) Placental diameter comparisons among WT, heterozygous and null placentas.

were significantly smaller at E11.5 (Figure 3.3D). One explanation for these data was that *Arid3a*-deficient placentas might be undergoing inflammatory-mediated swelling at E10.5 followed by apoptotic-mediated atrophy at E11.5 (further addressed below and in Figure 3.7A).

Arid3a mutant placentas have structural defects in placental organization

We first addressed abnormal size and weight mutant phenotypes by hematoxylin and eosin (H&E) staining. As shown in Figure 3.4, both E10.5 and E11.5 mutant placentas displayed abnormal organization of the junctional zone, disruption of both trophoblast giant cell (TGC) and spongiotrophoblast (SpT) layers, as well as an anomalous labyrinth layer. Consistent with these observations, immunohistochemistry (IHC) at E10.5 and E11.5 revealed that ARID3A protein was predominantly localized within nuclei of TGCs in WTs (Figure 3.4B). Additionally, we tested TGC markers at E9.5. Initially, TGCs were present in E9.5 *Arid3a* null placentas, but then are lost as placentas further developed (Figure 3.5A). Quantification indicated that the TGC densities within the junctional zone were also significantly reduced in *Arid3a* KO placentas (Figure 3.5B).

Next we examined the expression of HAND1, PROLIFERIN, and PLACENTAL LACTOGEN (PL1), which are required for the differentiation of cellular subtypes expressed in the TGC layer of the junctional zones (Scott et al., 2000)(Figure 3.4C-E). While HAND1 expression was evident in E10.5 WT placentas, mutants showed significantly reduced staining around the junctional zones (Figure 3.4C). Levels of PROLIFERIN and PL1 were highly disrupted in null placentas (Figure 3.4D,E). Consistent with disruption of the SpT layer, expression levels of SpT markers such as ASCL2 and CDX2, were significantly reduced at E11.5, as confirmed by RT-qPCR (Figure 3.4F,G; Figure 3.6B,C). These results indicate that Arid3a is required for the structural organization of various layers of the placenta.

Aberrant expression of subtype-specific markers in Arid3a null placentas

We further examined expression levels of key genes previously shown to be vital to placental development (Rossant and Cross, 2001). Upon KO of *Arid3a* (Figure 3.6A), genes required for trophoblast lineage development, SpT maintenance, labyrinthine development, and nutrient transport across the labyrinth pathway were highly down-regulated (Figure 3.6B,C). Conversely, key markers of ectoplacental cone function and chorioallantonic fusion were unchanged (Figure 3.6C).

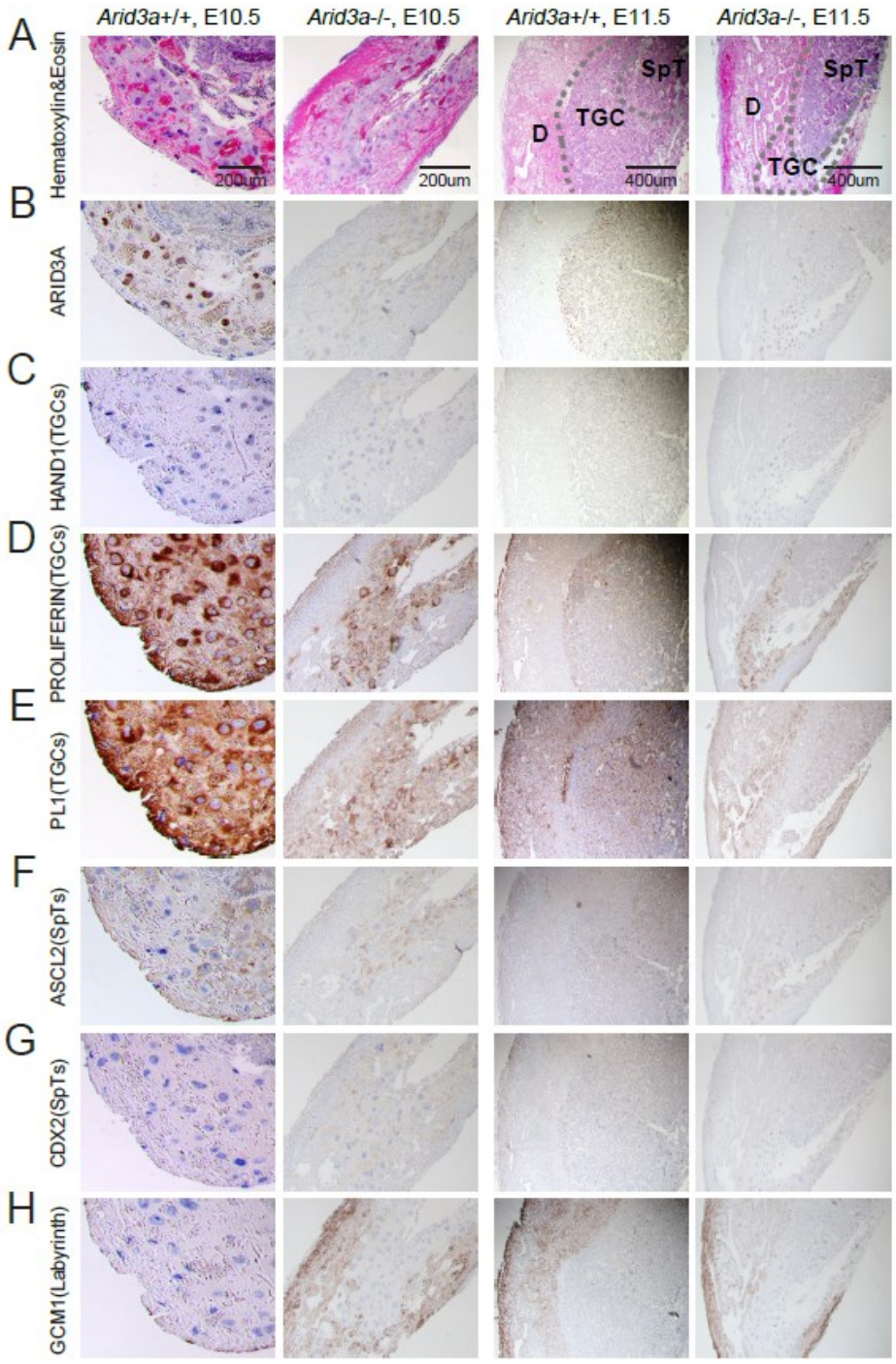


Figure 3.4. *Arid3a* KO in mouse embryos results in defects in placental structural organization. (A-H) Immunohistochemistry of *Arid3a* WT and KO sagittal placental cross sections stained with Hematoxyline and Eosin (H&E) and antibodies against: (A) ARID3A; (B) HAND1; (C) PROLIFERN; (D) PLACENTAL LACTOGEN (PL1); (E) ASCL2; (F) CDX2; (G) GCM1 (H). D, Decidua; TGC, Trophoblast giant cells; SpT, Spongiotrophoblasts

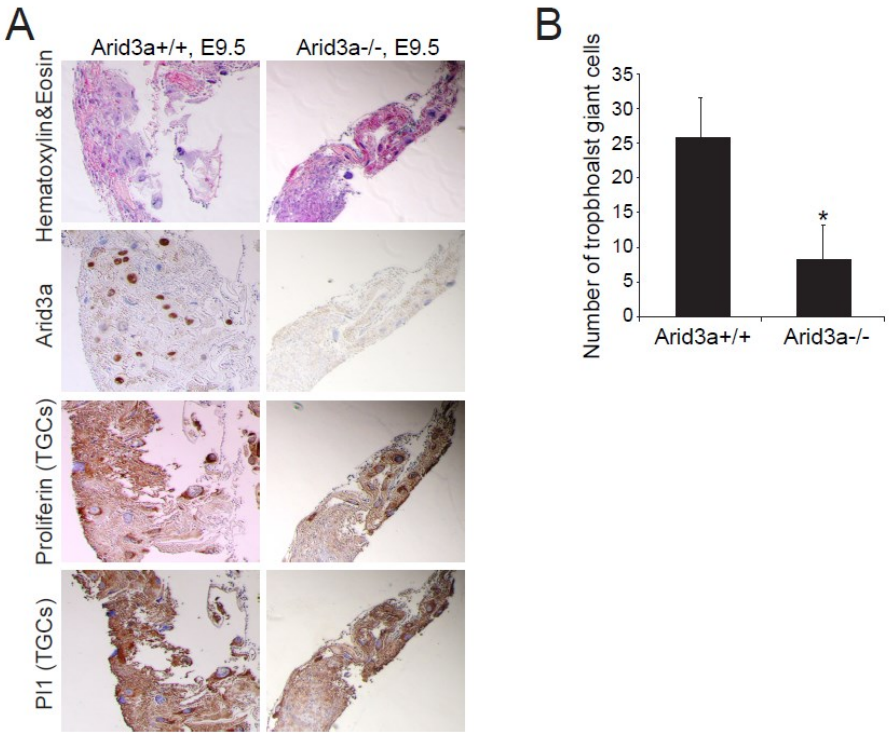


Figure 3.5. *Arid3a* KO in mouse embryos results in defects in placental structural organization. (A) Immunohistochemistry of E9.5 Arid3a WT and KO sagittal placental sections stained with Hematoxyline and Eosin (H&E) and antibodies against Arid3a, Proliferin, and Placental lactogen (Pl1). (B) Number of TGC counted per respective genotype at E10.5.

To confirm and extend the RT-qPCR results, we generated expression profiles of *Arid3a* KO and WT placentas by RNA-seq. We identified approximately 8,000 and 2,000 differentially expressed genes at E10.5 and E11.5, respectively (cut-off threshold ≥1.5-fold; Figure 3.6D,F). Differentially up-regulated genes at E10.5 were strongly enriched in gene ontology (GO) terms associated with immune system-related processes, including the response to external stimuli and defense response (Figure 3.6E). Conversely, genes down-regulated upon KO were enriched in general metabolic and structural processes, such as anatomical structure and pattern specification (Figure 3.6E).

Activation of the inflammatory response in *Arid3a* null placentas

Further inspection of our global expression profiles (Figure 3.6E,G) revealed that ARID3A is a critical regulator of innate immunity responses in the placenta. As shown in

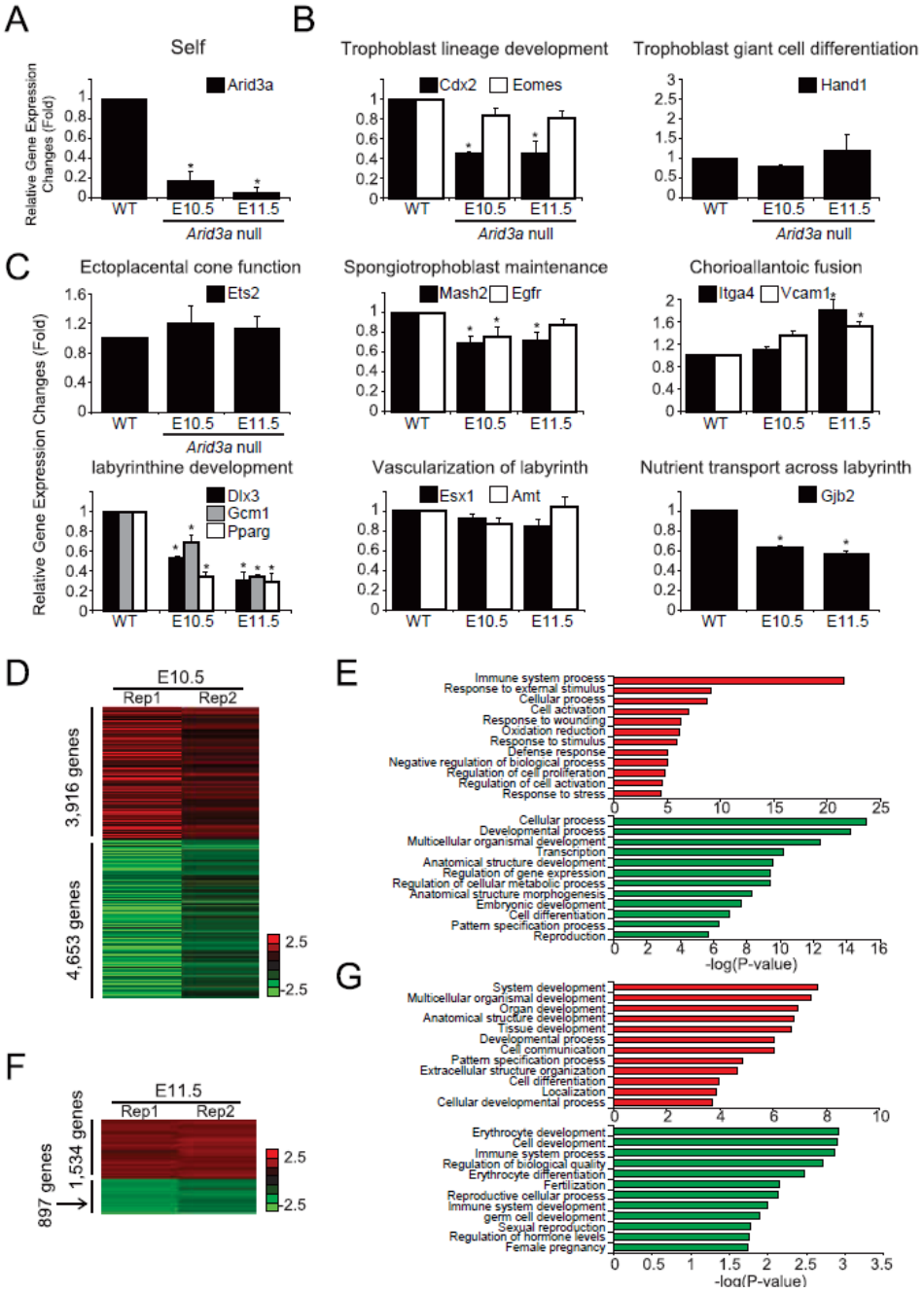


Figure 3.6. *Arid3a* KO results in aberrant expression of subtype-specific placental markers. (A) Expression levels of Arid3a in WT and KO placentas. (B,C) Expression levels of placental lineage markers as determined by RT-qPCR. All data are plotted relative to WT levels. Error bars indicate standard deviations of biological duplicates. (D, F) An unsupervised hierarchical clustering of transcript disruption by Arid3a KO using a cutoff threshold of 1.5-fold from expression profiles of Arid3 at E10.5 and E11.5, respectively. (E, G) Significantly enriched GO terms (biological functions) of differentially expressed genes upon KO of Arid3a at E10.5 and E11.5; red, up-regulated and green, down-regulated genes.

Figures 3.7A-C, KO of *Arid3a* at E10.5 led to up-regulation of transcripts encoding inflammatory chemokines (eg, *Ccl12, Ccl17, Ccl8, Ccl5*), cytokines (eg, *Il1a, Lta, Ltb, Mif*), and several other inflammatory mediators (eg, *Fcer1g, Nkx2-3*, and *Cd4*), whereas at E11.5, transcripts encoding pro-apoptotic factors (eg, *E2f3, Tnfrsf10b*, and *Bcl2l13*) were up-regulated. These results are consistent with the data of Figure 3.3D and our recent analyses of *Arid3a* KO hematopoiesis (Kim et al., 2016), suggesting that E10.5 *Arid3a*-deficient placentas are undergoing inflammatory-mediated swelling followed at E11.5 by apoptotic-mediated atrophy.

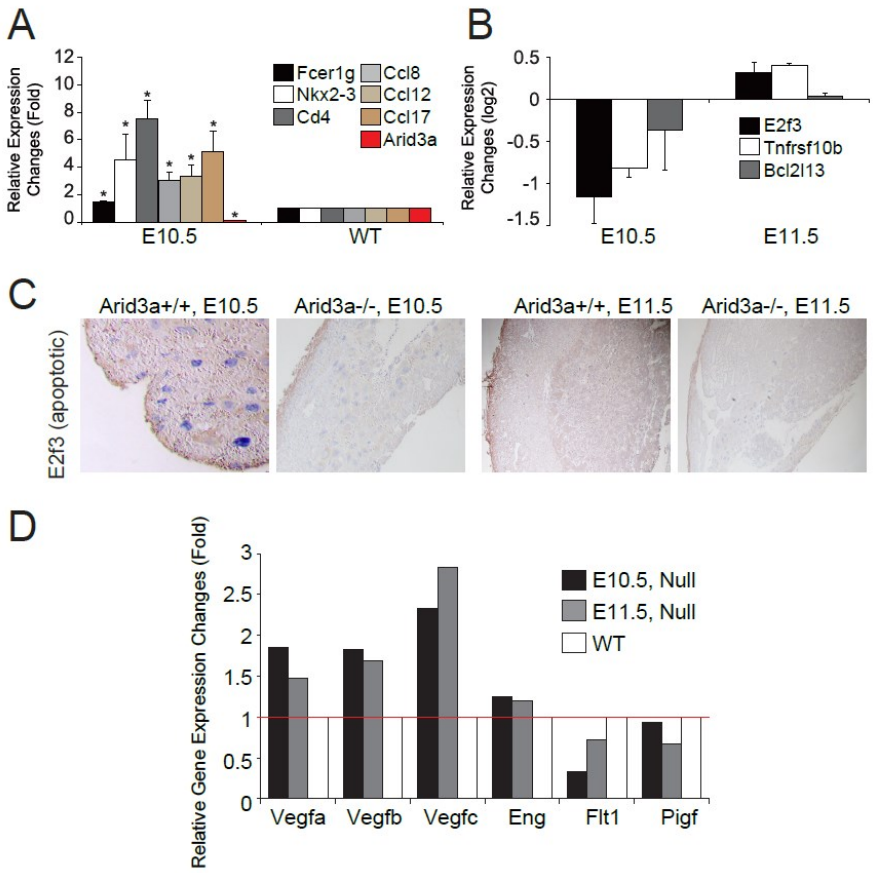


Figure 3.7. Activation of innate immune response in *Arid3a* null placentas. (A) Alteration of expression associated with innate immunity in *Arid3a* KO placentas assessed by RT-qPCR. (B) Expression levels of apoptosis mediators altered in *Arid3a* KO placentas as assessed by RNA-seq. (C) Immunohistochemistry of Arid3a WT and KO placental sections stained with antibodies against E2f3. (D) Expression levels of inflammatory response-related genes, cytokines and chemokines upregulated in *Arid3a* KO placentas as assessed by RNA-seq.

We also noted that several angiogenic factors and their receptors, including the soluble Fms-like tyrosine kinase (*sFlt-1*) and placental growth factor (*Plgf*) are significantly deregulated in *Arid3a* KO placentas at E10.5 but return close to normal levels at E11.5 (Figure 3.6F, 3.7D). Excessive inflammation and angiogenic imbalance often underlies symptoms associated with PE (Perez-Sepulveda et al., 2014); readdressed in Discussion).

Loss or gain of Arid3a disrupt expression levels of key TS and TGC differentiation markers

We first examined the effect of *Arid3a* loss on inflammatory chemokine expression following shRNA-based KD of *Arid3a*. We obtained > 70% KD efficiency at the mRNA level (Figure 3.8A). *Arid3a*-deficient TS cells revealed deregulation of several inflammatory chemokines (e.g., *Ccl8, Ccl12, Ccl17*) as well as upregulation of several inflammatory mediators (e.g., *Nkx2-3*, and *Cd4*) (Figure 3.8A). Next, we assessed expression levels of key markers of TE development. As shown in Figures 3.8B and C, well established markers (e.g., *Cdx2, Hand1, Gata3, Eomes*, and *Tcfap2c*) were down- or up-regulated upon *Arid3a* KD or OE, respectively.

Finally, to better understand the consequences of *Arid3a* loss on further differentiation to mature trophoblastic lineages, we performed *in vitro* differentiation of *Arid3a*-deficient TS cells as previously described (Tanaka et al., 1998)(Figure 3.8D). We observed marked deregulation of *Hand1*, which is essential for TGC differentiation (Scott et al., 2000), *Cdx2*, an early and essential marker of TE polarity and integrity of the TE epithelium (Strumpf et al., 2005), *Gcm1*, which promotes differentiation of underlying cytotrophoblast cells into the outer syncytiotrophoblast layer (Bainbridge et al., 2012) and *Ascl2*, which is required for the maintenance of TGC precursors (Guillemot et al., 1994). Also significantly deregulated was the labyrinth marker, *Pparγ*, which promotes labyrinthine trophoblast differentiation via Gcm1-regulation (Fournier et al., 2008).

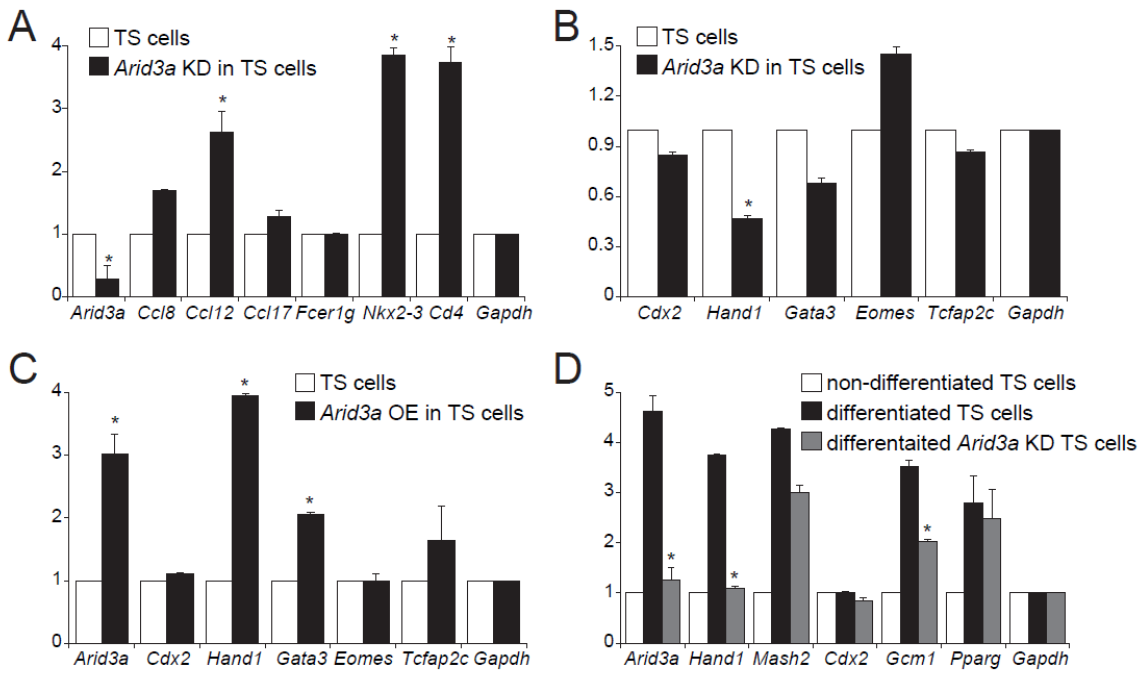


Figure 3.8. Loss or gain of *Arid3a* disrupts expression levels of key TE markers in mouse TS cells. (A) Inflammatory chemokine markers are up-regulated following *Arid3a* KD in mouse TS cells as measured by RT-qPCR. Asterisks indicate significance (minimally) of $p \leq 0.05$ of at least 3 independent biological replicate; error bars indicate standard deviation of biological triplicates. (B,C) Expression levels of key markers of TE development are down- or up-regulated upon *Arid3a* KD (B) or OE (C) in TS cells, respectively. (D) *In vitro* differentiation of Arid3a-deficient TS cells perturbs gene expression corresponding to essential layers of the placenta as compared to differentiated WT TS cells.

These results suggest that Arid3a gain or lost leads to significant changes in TS cell inflammation, development and differentiation, particularly within the TGC lineage, whose numbers and disorganization are apparent by IHC (Figure 3.4).

Human ARID3A levels are induced by BMP4 and ARID3A OE leads to down-regulation of pluripotency and up-regulation of TE-associated gene expression

Our meta-analyses of the data of Li et al (2013) indicated that *ARID3A* is highly expressed in the developing human placenta and in TS-like cells (Figure 3.2A,C). As an initial approach to circumvent the ethical limitations of studying human placental function *in vivo*, we generated human "TS-like" cells (Xu et al., 2002) by

addition of BMP4 to human iPS cells in feeder-conditioned medium (FCM) cultures. This treatment, while controversial (further addressed in Discussion), is generally acknowledged to induce TE differentiation.

BMP4 induction led to both increased ARID3A levels as well as cytoplasmic to nuclear translocation of ARID3A (Figure 3.9A,B). Transient transfection of *ARID3A* into these TS-like cells increased its levels ~8-fold (without alteration of the differentiated morphology induced by BMP4/FCM) and up-regulated the TE-markers, *GATA3* and *GCM1* (Figure 3.9A,C). Conversely, shRNA-mediated *ARID3A* KD de-repressed transcription of pluripotency genes (*OCT4* and *NANOG*) while repressing transcription of TE genes (Figure 3.9D). This may provide a mechanism by which ARID3A loss delays differentiation of human ES cells to TS-like cells.

3.4 DISCUSSION

Here we show that ARID3A plays a vital role in mouse placentation by contributing to the structural integrity of the placenta as well as by mediating communication among critical effectors, such as cells of the innate immune system. Our initial *in vitro* investigation in the human model further supports a requirement for ARID3A in initiation of ES to TS-like cell conversion through modulation of underlying transcriptional programs. Taken with our previous mechanistic observations (Rhee et al., 2014), our results identify ARID3A as an indispensable component of placental commitment, hemostasis and immune tolerance.

Anchoring the conceptus to the uterus is a critical placental function. In mice, TGCs perform this role, and the bHLH TF, Hand1, is critical (Riley et al., 1998; Scott et al., 2000). *Hand1* expression was not altered by loss of *Arid3a*. Instead, the most striking consequence of *Arid3a* deficiency was displayed morphologically. While specific sub-lineages can be delineated, they are highly disorganized with reduced cellularity (Figure 3.4). That along with the expression data of Figure 3.6 supports an interpretation in which both layers and genes underlying their formation are deregulated in *Arid 3a* nulls. Addressing this observation *in vitro* by employing isolated cell cultures would contribute

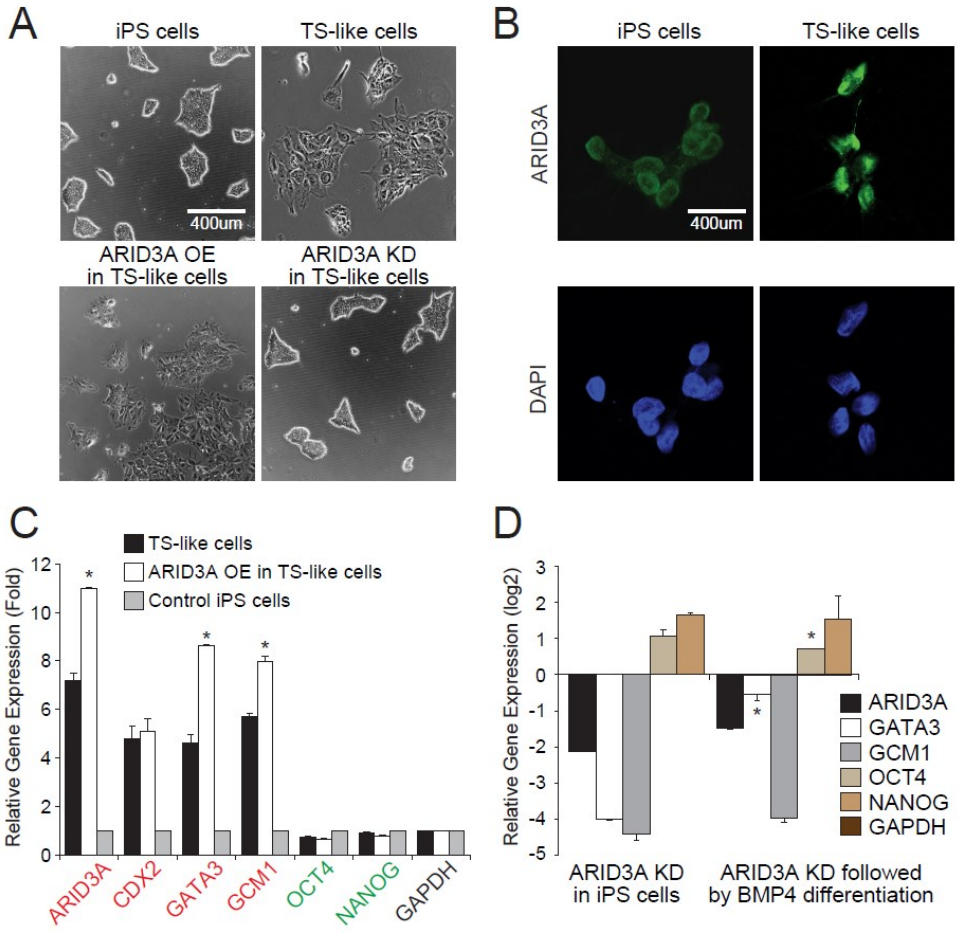


Figure 3.9. ARID3A loss or gain disrupts expression levels of key TE markers in human TS-like cells. (A) BMP4-mediated human TS-like cells with ARID3A OE show differentiated morphology. ARID3A KD in TS-like induces a rounded morphology, indicative of a pluripotent state indistinguishable from human iPS cells. (B) Immunofluorescence images of control iPS and TS-like cells stained with directly conjugated anti-ARID3A (green; Alexa 594) and DAPI (blue). (C) Trophoblast lineage markers (red) and pluripotency factors (green) are up-regulated following ARID3A OE as measured by RT-qPCR. Error bars indicate standard deviation of biological duplicates. (D) Trophoblast lineage markers and pluripotency factors are down-regulated following KD of ARID3A.

significantly, but it would not fully resolve the *in vivo* defects. That will require loss of function analyses of specific sub-lineage-restricted master transcription factors. Additional important questions to be resolved by future studies are the mechanisms by which Arid3a maintains this structural integrity and the precise consequences of Arid3a loss on TGC transcriptional programs.

Global RNA expression profiling confirmed that ARID3A function extends beyond the "first cell fate decision" (Rhee et al., 2014) to additional cellular processes vital to placental development, including placental structure, metabolism, immune tolerance and angiogenesis. Innate immune cells, typically placental NK, switch from a tolerogenic, anti-inflammatory phenotype to a cytotoxic, pro-inflammatory phenotype upon the sensing of pathogens or endogenous danger signals (Perez-Sepulveda et al., 2014). In our case, we suggest the trigger was loss of *Arid3a*. As cytotoxic effectors, innate immune cells create a state of inflammation, via cytokine and chemokine release, and placental ischemia, through reduced angiogenesis and increased vasoconstriction—the phenotype observed at E10.5 in *Arid3a*$^{-/-}$ placentas. We recently published a detailed analysis (Kim et al., 2016) of the inflammatory consequences of ARID3A loss on embryonic hematopoiesis. In those studies, conducted at the same embryonic time points as analyzed here, we observed inflammatory consequences of a prototypical T_H1 response—most notably, a significant increase in IFNα. However, the immune profiles of our null placentas displayed no significant overlap with the INFα target genes of Kim et al. (2016) nor with other targets deregulated in the null maternal embryo. This indicates that a large component of the observed placental inflammation is intrinsically derived. ARID3A is also required at later stages of embryogenesis for normal erythroid lineage differentiation and hematopoietic stem cell production (Kim et al., 2016; Webb et al., 2011).

The *in vivo* authenticity of the previously characterized mesoderm inducer, BMP4, in human TS induction has been challenged (Bernardo et al., 2011; Ezashi et al., 2012; Li and Parast, 2014). However, more recently and well after the BMP4 debate arose, a number of groups have confirmed that the BMP4 system can induce TS-like cell phenotype. For example, Kurek et al (2015) showed conclusively that BMP-alone targets are required and sufficient for mesoderm induction, whereas trophoblast induction is WNT dependent, suggesting that exclusive differentiation toward either lineage is possible in BMP4 cultures. Yabe et al (2016) found that, when BMP4 is used in combination with ACTIVIN and FGF2 signaling inhibitors, trophoblast differentiation was significantly more efficient and synchronous. Taken in this context, we feel that our

approach—to determine the effect of ARID3A loss or gain in producing *in vitro* "TS-like" cells from human ES/iPS cells—is rational.

We observed that, as in the mouse (Rhee et al., 2014), *Arid3a* overexpression under BMP4-mediated culture conditions did convert human iPS to a more TS-like phenotype via downregulation of pluripotency genes and upregulation of trophoblast genes (Figure 3.9C,D). However, not all our findings in mouse were conserved in the human. For example, in mouse blastocysts and placenta (Rhee et al., 2014), ARID3A activates CDX2, a transcription factor required for the initiation of TS commitment (Dietrich and Hiiragi, 2007). However, in the BMP4-induction hES cell system, ARID3A loss had no effect on CDX2 expression. This is consistent with observations that CDX2 expression is maximized in the non-trophoblast (mesoderm-derived) component of the human chorion (Bernardo et al., 2011; Niakan and Eggan, 2013). A further difference was that the reciprocal CDX2-OCT4 expression patterns established in the mouse are not conserved in the human (Hay et al., 2004). The human embryo shows a lag in trophoblast lineage segregation, with a period of time during which OCT4 and CDX2 are co-expressed in the TE (Niakan and Eggan, 2013). Nonetheless, we observed ARID3A-mediated repression of human *OCT4*, even though its human regulatory elements have diverged to a point such that they are not repressed when enforced into mouse TE (Molineris et al., 2011). While its specific role in human placentation remains to be elucidated, our findings establish that ARID3A is instrumental to trophoblast lineage determination.

Preeclampsia (PE) is a major cause of pregnancy-associated morbidity and mortality, affecting 2-5% of pregnancies worldwide after 20 weeks of gestation (Ananth et al., 2013). Normal placental function depends on trophoblastic invasion of the maternal decidua, myometrium, and blood vessels (Li and Parast, 2014). Fms-like tyrosine kinase 1 (sFLT-1) and placental growth factor (PlGF) are key among angiogenic and anti-angiogenic and mediators implicated in PE pathology (Levine et al., 2006). A recent study (Zeisler et al., 2016) demonstrated that the ratio of sFLT-1 to PlGF is elevated in the blood of pregnant women prior to the onset of PE. As shown in Figure 3.7, Arid3a activates both sFLT1 and PIGF, as well as the additional inflammatory agents,

ENDOGEN, and VEGFR-2. Additional Arid3a targets associated with PE include PPARγ and GCM1 (Chen et al., 2004; Waite et al., 2000). Our meta-analyses of placental PE data sets (Bilban et al., 2009; Founds et al., 2011) identified ARID3A and ELF5 as the only two deregulated transcription factors associated with human TE specification and/or differentiation.

In summary, ARID3A is vital not only for maintaining proper intrauterine growth but also for maintaining a properly balanced immune system during pregnancy. Our data indicate that, as suggested by previous TS cell studies *in vitro* (Rhee et al., 2014), ARID3A is required for normal murine trophoblast development *in vivo*. However, it is important to note that while numerous pathways are deregulated, the etiology of the initiating placental defect in the mouse remains to be determined. It is unclear and the focus of future studies as to whether the primary defect is a direct consequence of the *Arid3a* null placenta, is secondary to a failing embryo/fetus or is derived from a combination of these events. Finally, we suggest that the use of ARID3A as a biomarker may provide a diagnostic, noninvasive predictive molecule to identify mothers at risk for defective deep placentation syndromes.

CHAPTER 4: MECHANISMS OF TRANSCRIPTION FACTOR-MEDIATED DIRECT REPROGRAMMING OF MOUSE EMBRYONIC STEM CELLS TO TROPHOBLAST STEM-LIKE CELLS

4.1 INTRODUCTION

During normal embryonic development, differential gene expression across various tissues irreversibly determines cellular identity. However, advances in transcription factor (TF)-mediated direct reprogramming have revealed the plasticity of cell identity and the feasibility of cell fate conversion both *in vitro* and *in vivo* (Aoi et al., 2008; Davis et al., 1987; Feng et al., 2008; Ieda et al., 2010; Kulessa et al., 1995; Nerlov and Graf, 1998; Takahashi and Yamanaka, 2006). To successfully convert cell fate, global gene expression in conjunction with the chromatin landscape must be altered in the original cells to a state favorable to a reprogrammed cell type. However, little is known about the mechanism of how and to what extent altered expression of TFs modulates chromatin architecture and global gene expression program changes. Although models for gene activation during reprogramming have been proposed, little is known about how repression of genes specific for the original cell type occurs.

Induced pluripotent stem (iPS) cells can be generated from overexpression (OE) of a handful of TFs (e.g., Oct4, Sox2, Klf4, and Myc) in fibroblasts. Reprogramming to iPS cells can be broadly divided into two gene activation phases: a long stochastic phase followed by a shorter deterministic phase (Stadtfeld et al., 2008). Initially, cells undergo increased proliferation and alterations in histone modifications to initiate mesenchymal-to-epithelial transition. Later, cells stochastically undergo an epigenetic reset, as well as activation of their target cell-specific transcriptional networks to stabilize the

---

Reproduced in part from Catherine Rhee, Samuel Beck, Bum-Kyu Lee, Lucy LeBlanc, Haley O. Tucker, and Jonghwan Kim. Mechanisms of transcription factor-mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like cells. In preparation. (Contributions: C.R. and J.K. designed and analyzed the experiments. C.R., S.B., and B.L., performed data analysis. C.R. and J.K., wrote the manuscript with input from other authors)

reprogrammed cells. Another recently reported gene activation mechanism during reprogramming is the ability by which ectopically expressed TFs act as "pioneer" factors (Zaret and Carroll, 2011). Pioneer factors initially bind to closed chromatin of genes specific to the target cell type. Once bound, pioneer factors interact with chromatin modifying enzymes or remodeling complexes to convert closed into open chromatin, thereby activating initial cell-specific genes. For example, early in reprogramming, Oct4, Sox2, and Klf4 function as pioneer factors by accessing closed chromatin of distal regulatory elements of pluripotency genes, such as Esrrb and Sall4 (Soufi et al., 2015). Ascl1, a TF capable of converting fibroblasts to induced neuronal (iN) cells, also works as a pioneer factor (Wapinski et al., 2013). Although activation of repressed genes can indirectly suppress actively expressed genes, it remains unclear as to how active genes are directly repressed, and whether activation and repression of sets of genes occur simultaneously or sequentially in an ordered manner.

Embryonic stem (ES) cells can undergo trans-differentiation or reprogramming to trophoblast stem (TS)-like cells whose properties are highly similar to genuine multipotent TS cells (Kuckenberg et al., 2010; Niwa et al., 2005; Ralston et al., 2010; Rhee et al., 2014). Several studies showed that the introduction of individual trophectoderm (TE)/TS cell-specific TFs (including Cdx2, Eomes, Tcfap2c, Gata3, and Arid3a) into ES cells drives their fate to TS-like cells. This transition is evidenced by alterations in cell morphology and in global gene expression in a manner similar to genuine TS cells (Kuckenberg et al., 2010; Niwa et al., 2005; Ralston et al., 2010; Rhee et al., 2014). In particular, TS-like cells generated by OE of Cdx2 and Arid3a were successfully incorporated into the TE of developing embryos and contributed to placental lineages *ex vivo* (Rhee et al., 2014). These findings raised the feasibility of generating functional TS-like cells from ES cells. ES to TS-like cellular conversion also has served as a valuable *in vitro* system for mechanistic understanding of the "first cell fate decision" (Zernicka-Goetz, 2004) as well as for TE lineage specification and differentiation. Thus, we reasoned that such an approach would allow us to thoroughly

interrogate mechanisms of transcriptional and epigenetic regulation by OE of single TFs during reprogramming.

Here, we employed an ES to TS-like cell reprogramming system via OE of three key TE/TS cell-specific TFs – Cdx2, Arid3a, and Gata3 (herein referred to as CAG factors). Each of these TFs are well-known for being instrumental in trophoblast differentiation and placental development (Niwa et al., 2005; Ralston et al., 2010; Rhee et al., 2014). For instance, OE of Cdx2 and Arid3a in ES cells can generate functional TS-like cells *ex vivo,* and Gata3 is one of the factors used to generate induced TS (iTS) cells from fibroblasts. We investigated the dynamics of CAG factor binding as well as subsequent effects on chromatin accessibility and global gene expression during the initiation phases of reprogramming. We found that CAG factors orchestrate reprogramming of ES cells to TS-like cells via a two-step mechanism. First, each of these factors represses ES cell-specific genes through decommissioning of their associated active enhancers in ES cells. Second, individual CAG factors activate TS cell-specific genes by switching their own DNA-binding motifs from one that binds to ES cell-specific regulatory regions to elements specific to TS cells. Ultimately, each CAG factor gradually changes its chromosomal targets throughout the reprogramming process to achieve sequential repression followed by activation.

4.2 MATERIALS AND METHODS

Cell culture

Mouse J1 ES cells were cultured in DMEM (Dulbecco's modified Eagle's medium) supplemented with 18% fetal bovine serum (FBS), 2mM L-glutamine, 100 μM of non-essential amino acid supplement, nucleoside mix (100X stock, Sigma), 100 μM of β-mercaptoethanol, 1000U/ml of recombinant leukemia inhibitory factor (LIF, Chemicon), and 50U/ml of penicillin/streptomycin. ES cells were plated on 0.1% gelatin coated dishes. Mouse TS cells were maintained at a ratio of 3:7 of TS medium to MEF-conditioned TS medium, supplemented with 25ng/ml Fgf4, and 1μg/ml heparin. TS medium is RPMI 1640 (Roswell Park Memorial Institute medium, Gibco) supplemented

with 20% FBS, 100μM β-mercaptoethanol, 2mM L-glutamine, 1mM sodium pyruvate, penicillin (50U/ml), and streptomycin (50mg/ml). MEF-conditioned medium is TS medium conditioned by MEF. MEF were treated with mitomycin, followed by culturing for 3 days. The medium was collected every 3 days for three times. 293T cells were maintained in DMEM supplemented with 10% FBS, 2mM L-glutamine, and 50U/ml of penicillin/streptomycin. All cells were incubated at 37°C, 5% CO_2.

Generation of stable cell lines

Individual coding sequences of CAG factors were amplified by PCR using mouse cDNA clones (Origene, MC208274 for Cdx2, MC205205 for Arid3a, and mouse TS cell cDNA for Gata3) as a template, and then cloned into the pEF1α-FLBIO vector (Wang et al., 2006). The vector was transfected into BirA expressing ES cells by electroporation. The positive biotin-tagged OE stable cells were picked and maintained in ES media supplemented with puromycin (Invitrogen) and geneticin (Gibco). Ectopic expression levels were measured by RT-qPCR and Western blotting. Primer sequences for PCR are listed in Appendix.

Generation of inducible cell lines

Coding sequence of Arid3a amplified by PCR using mouse cDNA clones (Origene, MC205205) as a template, and then cloned into the pLVX-TRE3G-ZsGreen1 vector (Clontech, 631361). Clones were sequence-verified prior to use. Lentiviral production and infection were performed as described below. ES cells were plated at ~1 x 10^6 cells per 6-well plate with virus-containing supernatant (co-infection of pLVX-TRE3G-ZsGreen1 and pLVX-EF1α-Tet3G vectors). The cells were placed under neomycin and geneticin selection for 2 days followed by replacement of medium with fresh ES cell medium plus 1μg/mL doxycycline. ES cell medium containing doxycycline was changed each day for ~6 days. Ectopic expression levels were measured by RT-qPCR. Primer sequences for PCR are listed in Appendix.

Lentiviral production and infection

Lentiviral production was performed in 293T cells. 293T cells were plated at ~6 X 10^6 cells per 100mm dish and incubated overnight. Cells were transfected with 6μg of pLVX-TRE3G-ZsGreen1 or pLVX-EF1α-Tet3G vectors (for the inducible cell line, Clontech, 631359) with 4μg of pCMV-Δ8.9 and 2μg of VSVG helper plasmid using Fugene (Promega), according to the manufacturer's instructions. After 15 hours, 293T medium was replaced with ES medium. The supernatants containing viral particles were harvested 48 hours after transfection, filtered through 0.45μm pore-size cellulose acetate filters, and supplemented with polybrene (Millipore). Infections were performed with cells plated at a density of ~1 x 10^6 cells per 6-well plate with virus-containing supernatant supplemented with polybrene (Millipore).

Quantitative gene expression analysis

For RT-qPCR, total RNA was first isolated using the RNeasy Mini Kit (Qiagen). 500ng total RNA was then used to synthesize cDNA with qScript cDNA supermix (Quanta). RT-qPCRs were performed using PerfeCTa SYBR Green FastMix (Quanta) with 2uL of 20X diluted cDNA generated from total RNA. Using Primer 3 (Koressaar and Remm, 2007), RT-qPCR primers were designed to amplify an ~100bp region containing the junction between the two exons. For ChIP-qPCR, the primers were designed to amplify ~100bp regions centered on the putative binding sites. The Ct values obtained from qPCR were normalized against *Gapdh* for gene expression and against *Gfi1b*_proximal sites to determine relative gene expression and ChIP enrichment, respectively. Primer sequences for qPCR are listed in Appendix.

Western blotting

Proteins were lysed from cultured cells using RIPA Lysis and Extraction Buffer (G-Biosciences) with protease inhibitor cocktail (Roche), followed by incubation at 100°C for 20 minutes. Cell lysates were separated by electrophoresis on 4-20% gradient acrylamide gels and transferred to PVDF membranes. The blots were blocked with TBS-T (20mM Tris-HCl, pH 7.6, 136mM NaCl, and 0.1% Tween-20) containing 5% BSA (bovine serum albumin) for 1 hour and then incubated with primary antibody solution at

4°C overnight. After washing with TBS-T, the membranes were incubated with HRP-conjugated secondary antibodies for 1 hour at room temperature (RT). Then the membrane was washed with TBS-T, followed by developing with ECL reagents (GE Healthcare). Antibodies used for Western blotting were anti-Cdx2 (1:1000, ab88129, Abcam), anti-Arid3a (1:5000;(Webb et al., 2011)), anti-Gata3 (1:1000, SC-9009, Santa Cruz), anti-β-actin (1:10000, ab20272, Abcam), and anti-Hdac1 (1:1000, ab7028, Abcam).

Co-immunoprecipitation

One-step affinity purification with streptavidin-agarose beads (Invitrogen) using ~2 x 10^7 cells from ES cell lines expressing BirA only or BirA plus biotin-tagged proteins was performed as previously described (Kim et al., 2010). The final pulled-down proteins were eluted in Laemmli buffer prior to Western blotting.

RNA-sequencing analysis

Libraries were prepared with an RNA library preparation kit (E7490, NEB) using 1µg of RNA obtained by the RNeasy Mini Kit (Qiagen). RNA-seq libraries were sequenced with a 1 x 75-bp strand-specific protocol on a NextSeq 500 (Illumina). Data were analyzed using a high-throughput next-generation sequencing analysis pipeline: FASTQ files were aligned to the mouse genome (mm9, NCBI Build 37) using TopHat2 (Kim et al., 2013). Gene expression for the individual samples was calculated with Cufflinks (Trapnell et al., 2010) as RPKM values. RNA-Sequencing data are submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE90752.

ChIP-sequencing

ChIP assays were performed with ES cell lines expressing BirA only (reference) or both BirA and biotin-tagged proteins (samples) as previously described (Kim et al., 2008) using streptavidin magnetic particles (Roche). Additional ChIP assays were carried out in ES, OEArid3a_ES+, OEArid3a_TS+, and TS cells using Arid3a (Webb et al., 2011), Hdac1 (ab7028, Abcam), and H3K27ac (ab4729-100, Abcam) antibodies. ChIP-

sequencing libraries were generated using ChIP-seq library prep kits (NEB) according to the manufacturer's instructions. ChIP-sequencing libraries were sequenced using Illumina NextSeq 500 at the Genomic Sequencing and Analysis Facility (GSAF), The University of Texas at Austin. ChIP-Sequencing data are submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE90752.

ATAC-sequencing

ATAC assays were performed as previously described (Buenrostro et al., 2015). We used ~ 50,000 cells to start the experiment. The cells were incubated with transposition reaction mix for 30 min followed by PCR reaction for 18 cycles. ATAC samples in the ~250bp range were isolated using E-gel size select chromatography. ATAC-sequencing was performed using Illumina NextSeq 500. ATAC-Sequencing data are submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE90752.

Identification of ChIP- and ATAC-sequencing peaks

FASTQ files were aligned to the mouse genome (mm9, NCBI Build 37) using Bowtie 2.2.5. For the identification of peaks, model-based analysis with the ChIP-seq (MACS 2.1.1) peak caller (Zhang et al., 2008) was used with a default setting (macs2 callpeak –t Sample.sam –c Control.sam –f SAM –g mm –n) with keeping 5 duplicates. The ChIP-seq signals resulting from BirA-only expressing ES cells and Mock samples were subtracted to remove background noise for ChIP by incubation with streptavidin magnetic particles and native antibodies, respectively. ATAC-seq signals were normalized to the default background. We readily identified several thousand, significant individual CAG occupied loci upon filtering high-quality peaks based on p-values. Genes occupied by TFs within a 10Kb window (8Kb upstream to 2Kb downstream of TSS) were considered target genes and used for subsequent analyses. In order to summarize the peak signal enrichment over control experiments, z-score bedGraph files were produced, and following background subtraction, the bedGraph files were constructed using MACS2 version 2.1.1 with 'bdgcmp –m logLR' command. Peaks were visualized using the Integrative Genome Viewer from the Broad Institute.

Motif analysis

Both de novo and known motifs in peak regions in the genome were identified using Homer (Heinz S et al., 2010) with the fragment size for motif finding set as 100bp.

Heat maps

All heat maps were generated using JavaTreeview (Saldanha, 2004).

Gene Ontology (GO) analysis

DAVID 6.7 (Huang et al., 2009) was used for analysis of differentially expressed genes obtained from the expression profile data. TF targets and Genomic Regions Enrichment of Annotation Tool (GREAT) (McLean et al., 2010) was used for ChIP-seq peak analysis.

Correlation maps

Common binding sites of the indicated TFs were identified by peak calling with an overlap analysis. The binding sites of two TFs were calculated for with a paired-end Pearson correlation coefficient.

Principle component analysis (PCA)

The R prcomp was used to perform PCA on the RNA-seq data.

Dendrogram

R hcluster was used to generate dendrograms of RNA- and ATAC-seq data.

Public data sets used

ChIP-sequencing data sets used for the analyses were obtained from the Gene Expression Omnibus (GEO) database under the accession numbers GSE31039 (H3K27ac, ES cells), GSE42207 (H3K27ac, TS cells), and GSE11724 (Oct4, ES cells).

4.3 RESULTS

Ectopic expression of individual CAG factors in ES cells promotes changes in global gene expression and chromatin landscape toward TS cells

OE of either of the aforementioned TE/TS cell-specific CAG factors (Cdx2, Arid3a, and Gata3) previously was shown to convert mouse ES cells to TS-like cells (Niwa et al., 2005; Ralston et al., 2010; Rhee et al., 2014) To determine how these CAG factors initiate reprogramming, we designed our experimental set-up as shown in Figure 4.1A. Individual CAG factors were expressed under the control of constitutively active pEF1α promoters. After establishing clones, the cells were maintained for 4 days in ES cell culture media (ES+). The ES media was then replaced with TS media (TS+) to optimize conditions for reprogramming (detailed procedures in Materials and Methods). During the ES to TS-like cell fate conversion, we monitored changes in cellular morphology, global gene expression, chromatin accessibility, and CAG factor occupancy at days 4 (ES+) and 8 (TS+).

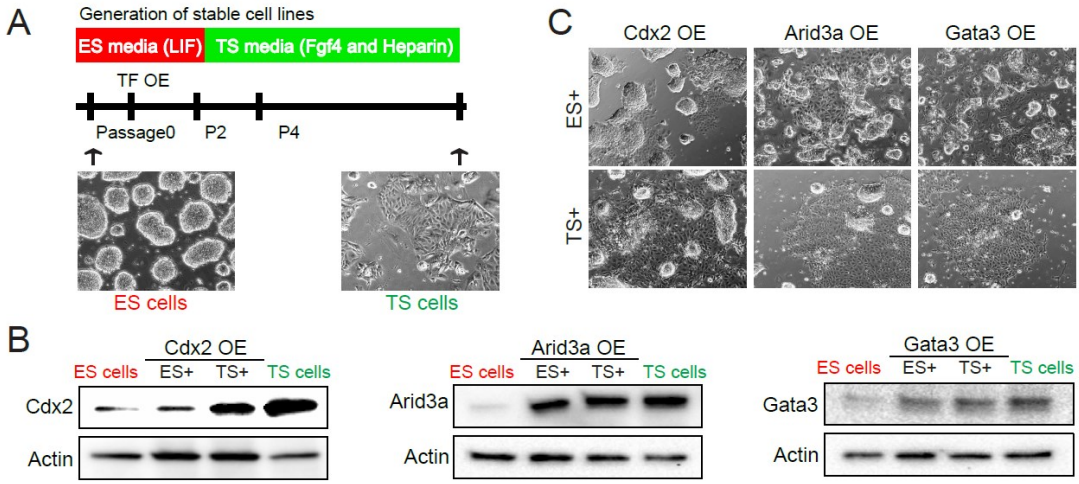


Figure 4.1. OE of single CAG factors in ES cells converts ES cells to TS-like cells. (A) Schematic diagram of the experimental design employed to generate TS-like cells from ES cells. (B) Western blotting showing protein levels of the CAG factors in ES cells and TS cells at different stages (ES+ and TS+) of the reprogramming upon OE of CAG factors. β-Actin was used as a loading control. (C) Bright field images depicting changes in cell morphologies associated with conversion of ES to TS-like cells at different stages of reprogramming.

We first confirmed OE of CAG factors in ES cells by Western blotting (Figure 4.1B,C). Consistent with previous reports (Niwa et al., 2005; Ralston et al., 2010; Rhee et al., 2014), OE of individual CAG factors in ES cells promoted dramatic morphological changes. Cells transitioned from spheroid colony morphology, typical of undifferentiated ES cells, to a flattened, epithelial-like morphology, indicative of an early stage of cellular conversion; this occurred even before replacing ES+ media with TS+ media (Figure 4.1C; upper panels). Under TS+ media conditions, cells adopted an even more prominent TS-like morphology (Figure 4.1C; lower panels). Parallel to these morphological changes, TE/TS cell-specific marker genes also were upregulated over time (Figure 4.2A), indicating that our experimental system was valid for investigating the early stages underlying CAG-mediated conversion of ES to TS-like cells.

Next we determined to what degree CAG factors affect global gene expression and chromatin configuration during ES to TS-like reprogramming, we profiled global gene expression by RNA-seq and chromatin accessibility by ATAC-seq at this early stage of reprogramming and then compared the results to transcriptional and chromatin analysis in ES and TS cells as controls. Indeed, at this early stage of reprogramming, global mRNA expression (and to a lesser extent, chromatin accessibility) had begun to transition in the OE cell lines from an ES cell state to a TS-like state. As shown in Figure 4.2B, we observed reduced levels of key ES cell-specific markers (including Oct4, Sall4, and Stat3) and induced levels of TE/TS cell-specific markers (including Eomes, Krt8, and Tead4) in CAG factor OE cells as compared to untreated ES cells. Principal component analysis (PCA) of these expression profiles showed a significant shift in the global expression pattern of ES cells toward TS cells (Figure 4.2C). Notable among the CAG factors, Arid3a induced an expression pattern most similar to that of bona fide TS cells (Figure 4.2D). Consistent with our RNA-seq results, OE of individual CAG factors induced alterations in ES cell chromatin accessibility more similar to that of TS cells (Figure 4.3A,B). Global changes in chromatin accessibility were minor compared to

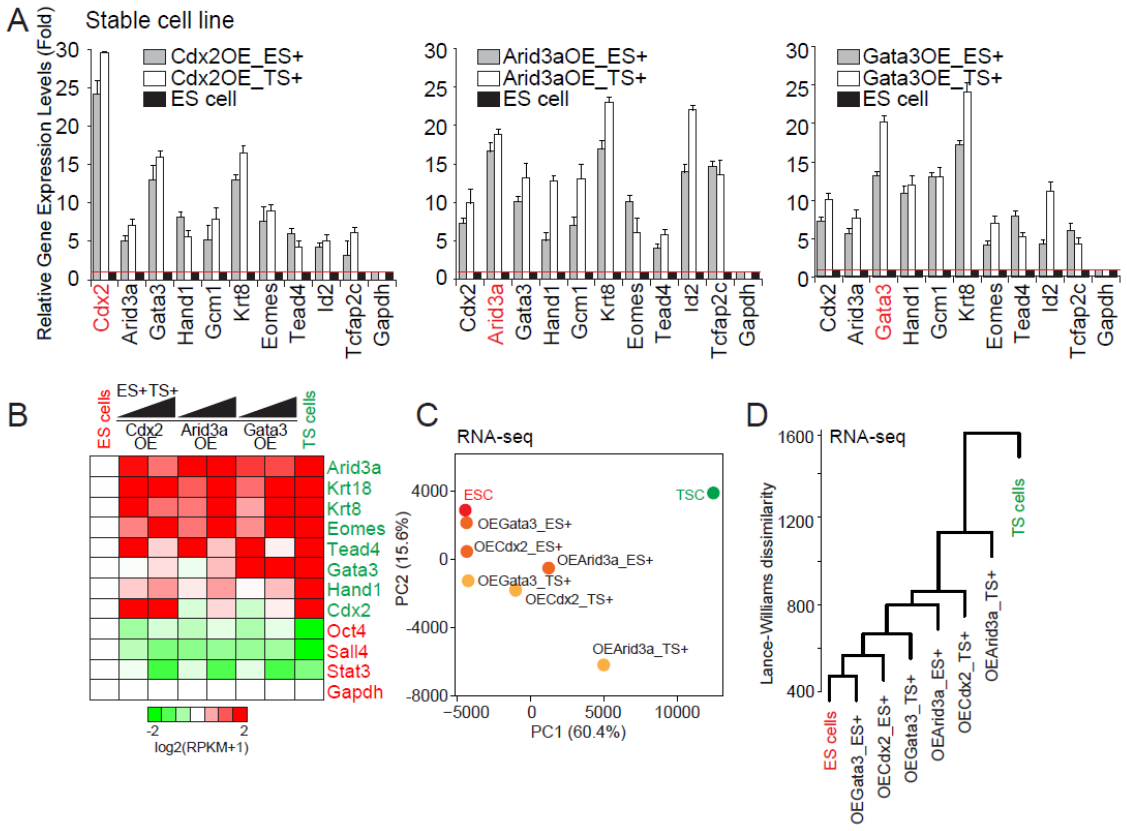


Figure 4.2. Ectopic expression of individual CAG factors in ES cells promotes transition of global expression toward TS cells. (A) Expression analysis of TS cell-specific genes by RT-qPCR in stable individual CAG-OE ES cells. (B) A heatmap showing relative gene expression levels of ES- and TS cell-specific genes in reprogrammed cells relative to ES cells upon OE of CAG factors. ES cell- and TS cell-specific genes are labeled in red and green to the right side of the heatmap, respectively. (C) Principal component analysis (PCA) of time-course RNA-seq data showing gradual transition of the transcriptome from ES cells toward TS-like cells. (D) A dendrogram showing a similarity of expression profile among ES, TS, and CAG-mediated reprogrammed cells.

to changes in expression levels. Still, a small portion of these changes are involved to be crucial markers of successful ES cell to TS-like cell conversion. Gene ontology (GO) term analysis revealed that regions that became more accessible (open) were enriched in genes involved in TE and placental development, while regions that became more inaccessible (closed) during reprogramming were implicated in embryonic stem cell maintenance (data not shown). Collectively, these results indicated that individual expression of CAG factors initiates

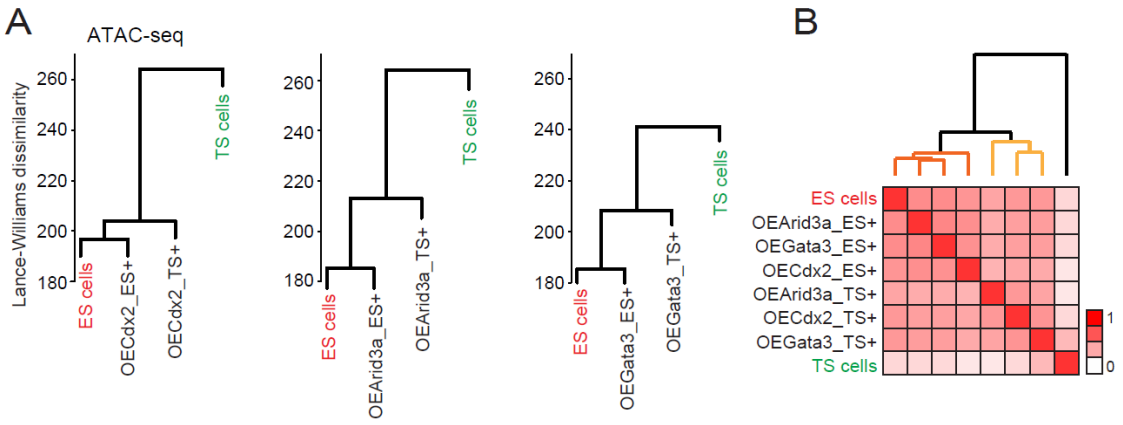


Figure 4.3. Ectopic expression of each CAG factors in ES cells leads to changes in chromatin landscapes toward TS cells. (A) Dendrograms showing a similarity of chromatin landscapes among ES, TS, and CAG-mediated reprogrammed cells. (B) A heatmap depicting similarity of chromatin openness among ES, TS, and different stages of reprogrammed cells.

ES to TS-like cellular conversion by not only modulating global gene expression, but also by reshaping the chromatin landscape.

CAG factors directly regulate both ES and TS cell-specific genes during ES to TS-like cell reprogramming

To examine how OE of CAG factors mediate global transcriptional changes during the reprogramming, we determined which classes of genes were directly regulated by CAG factor binding via genome-wide ChIP-seq. We found several thousand statistically significant individual CAG factor binding sites (detailed in Materials and Methods). Previous studies had shown that cellular identity is largely determined by distal regulatory elements, especially enhancers that are targets of cell type-specific TFs, during both development and reprogramming (Heinz et al., 2015). Consistent with this, we found that >60% of CAG target sites are within regions distal to their respective transcriptional start sites (TSSs), primarily distributed within intergenic (~40%) and upstream (~20%) regions during the initial stage (day 4, passage 2) of reprogramming (Figure 4.5A).

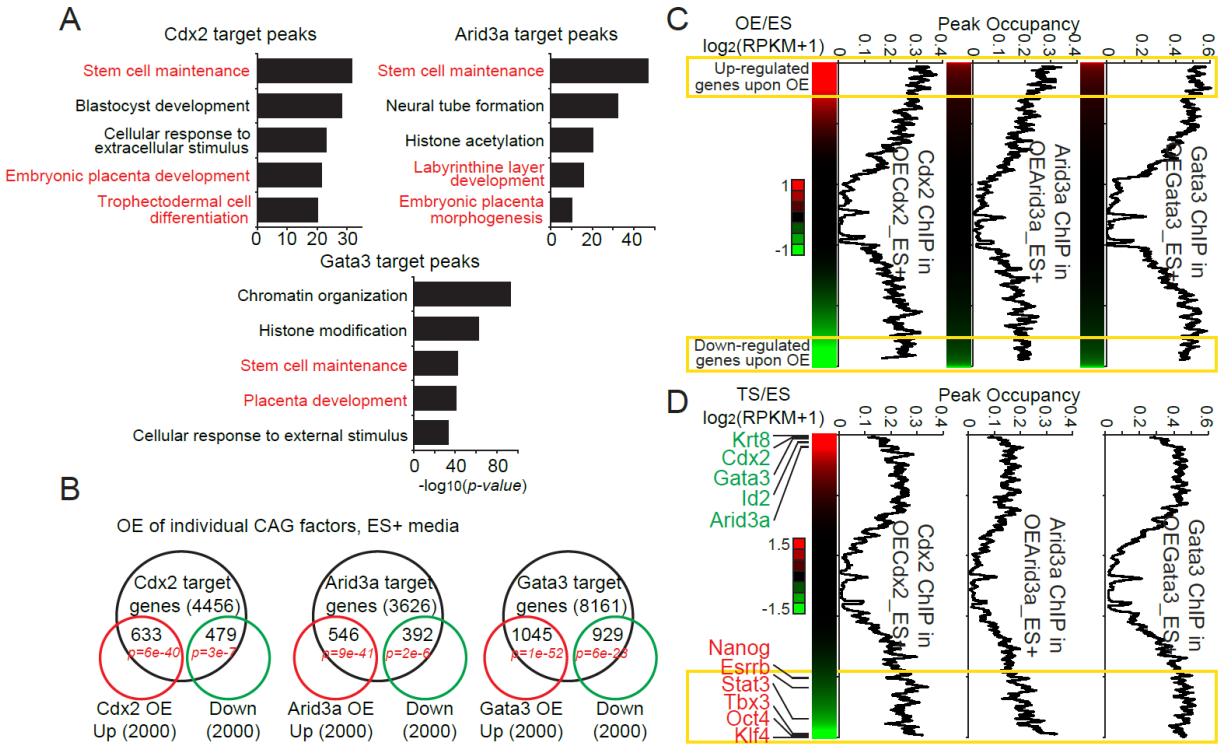


Figure 4.4. CAG factors possess dual functions during the reprogramming (earlier stage). (A) Bar graphs showing enriched GO terms of individual CAG factor targets. ES- and TS cell-specific functions are highlighted in red. (B) Venn diagrams depicting overlaps among genes bound by each CAG factor and genes that are either up- or down-regulated in ES cells upon OE of individual CAG factors. *p-values* were calculated using hypergeometric tests. (C-D) Heatmaps ranking relative gene expression of CAG OE cells relative to ES cells (C) and TS cells relative to ES cells (D). Averaged occupancy scores calculated by moving average (window size, 250; bin size, 1) for individual CAG factors are plotted in black lines.

We next performed GO analysis of the genes associated with chromosomal targets of each CAG factor. As shown in Figure 4.4A, CAG factor targets were largely associated with both ES cell-specific terms (e.g., stem cell maintenance and blastocyst development) and TS cell-specific terms (e.g., embryonic placenta development and trophectodermal cell differentiation). Thus, individual CAG factors act as distal enhancer binding proteins during reprogramming to directly regulate both ES and TS cell-specific genes. Comparison of our ChIP-seq and RNA-seq data revealed that both activated and repressed genes upon OE of each TFs were largely direct targets of CAG factors during

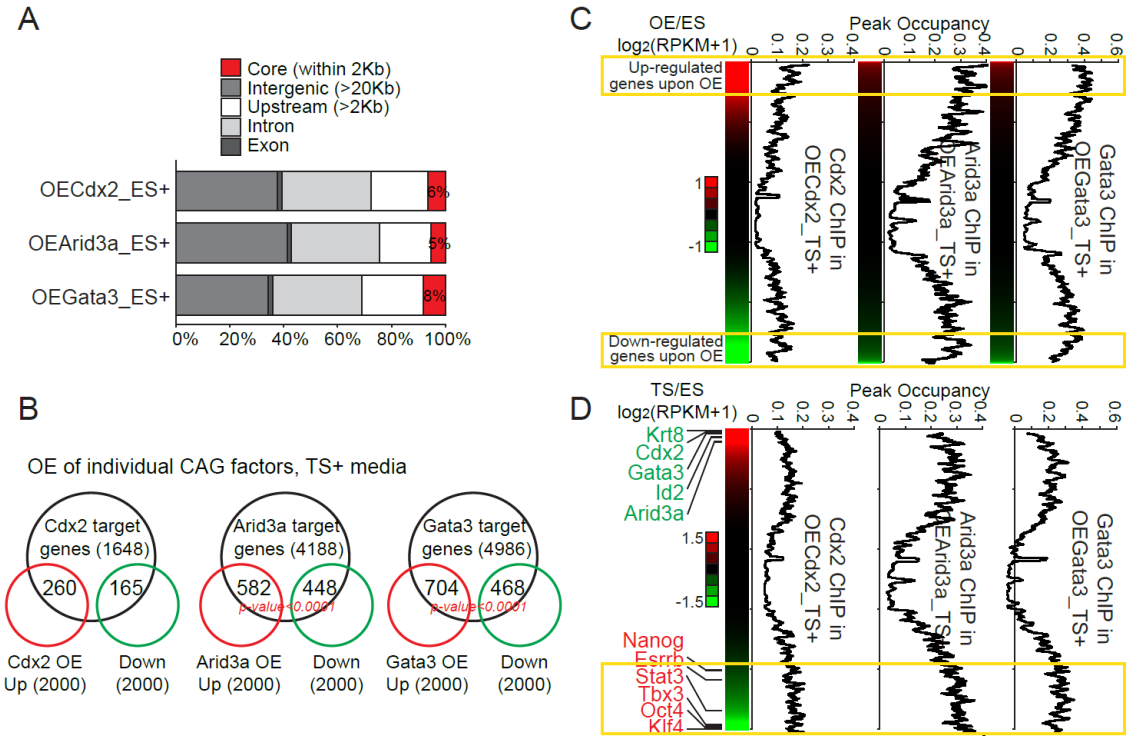


Figure 4.5. CAG factors possess dual functions during the reprogramming (later stage). (A) Bar graphs showing distribution of the binding sites of each CAG factors across the genome. (B-D) The same analyses shown in Figure 4.4B,C,D in order in TS+ media instead of ES+ media.

the early phase of reprogramming (Figure 4.4B, 4.5B). These results indicated that each CAG factor displays dual functions during reprogramming.

To further characterize the relationship between CAG factor binding and differentially regulated genes, we analyzed the ChIP-seq data as a moving window average (window size, 250; bin size, 1; Figure 4.4C, 4.5C). We found that stronger peaks, which indicate stronger occupancy of the factor bound to DNA, are generally correlated with greater changes in gene expression (Figure 4.4C, 4.5C). Then, to determine which classes of genes are directly regulated by CAG factors, we compared the occupancy of target genes upon OE of individual CAG factors in ES cells to their relative expression TS (Figure 4.4D, 4.5D). We found that the genes that are highly expressed in ES cells, well-established ES cell-specific genes, are, in fact, direct targets of CAG factor binding, whereas TS cell-specific TFs are less strongly

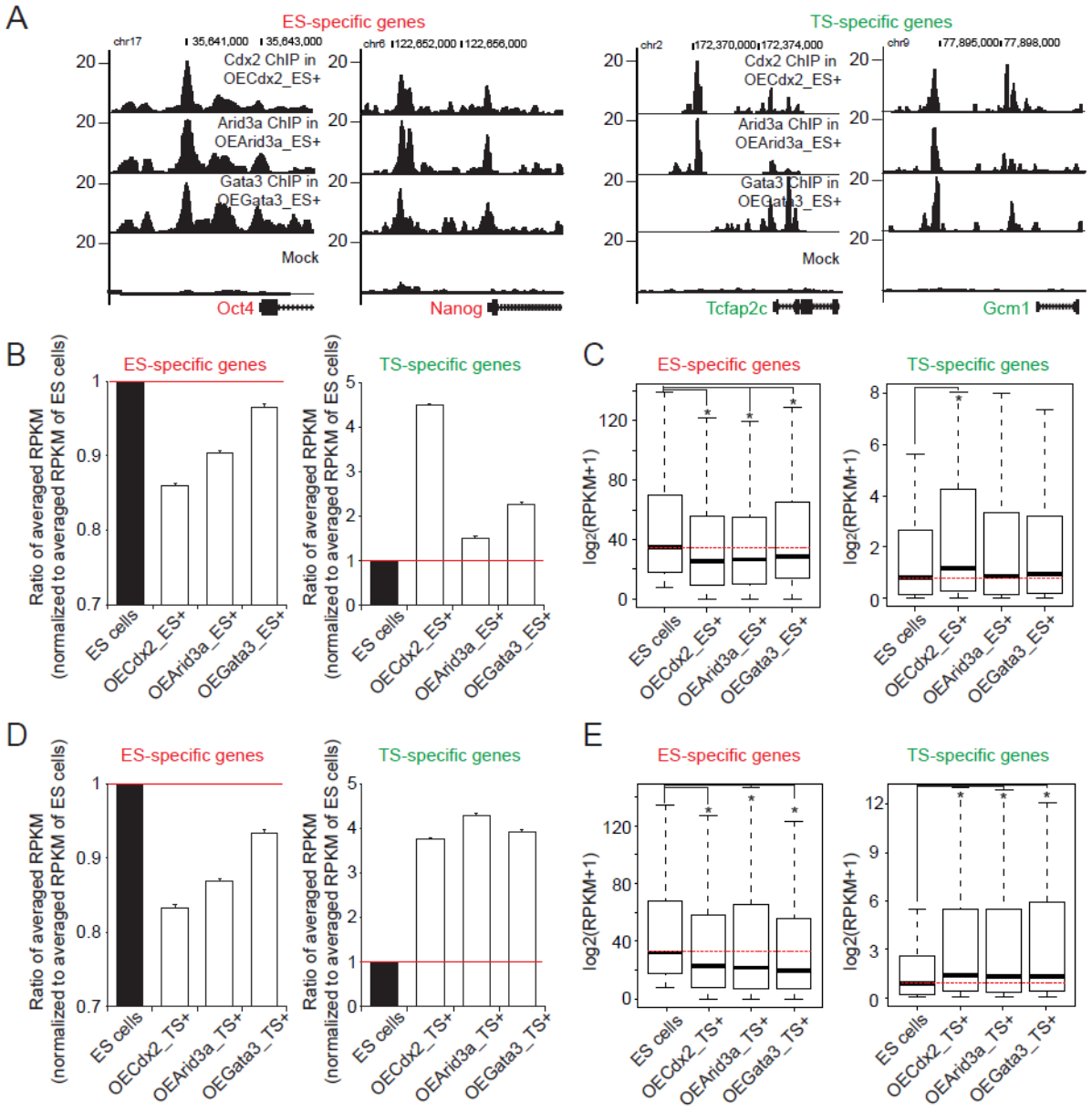


Figure 4.6. CAG factors directly activate and repress TS- and ES cell-specific genes, respectively. (A) Signal track images depicting occupancy of each CAG factor at both ES- and TS cell-specific genes upon OE in ES cells. (B) Bar graphs showing average expression levels of the top 150 ES cell-specific genes and the top 150 TS cell-specific genes upon OE of each CAG factor in ES cells. (C) Boxplots showing the distribution of expression levels of ES- and TS cell-specific genes upon OE of individual CAG factors in ES cells. The red dotted lines indicate median expression of ES- or TS cell-specific genes in ES cells. (D,E) The same analyses shown in Figure 4.6B,C in order in TS+ media instead of ES+ media.

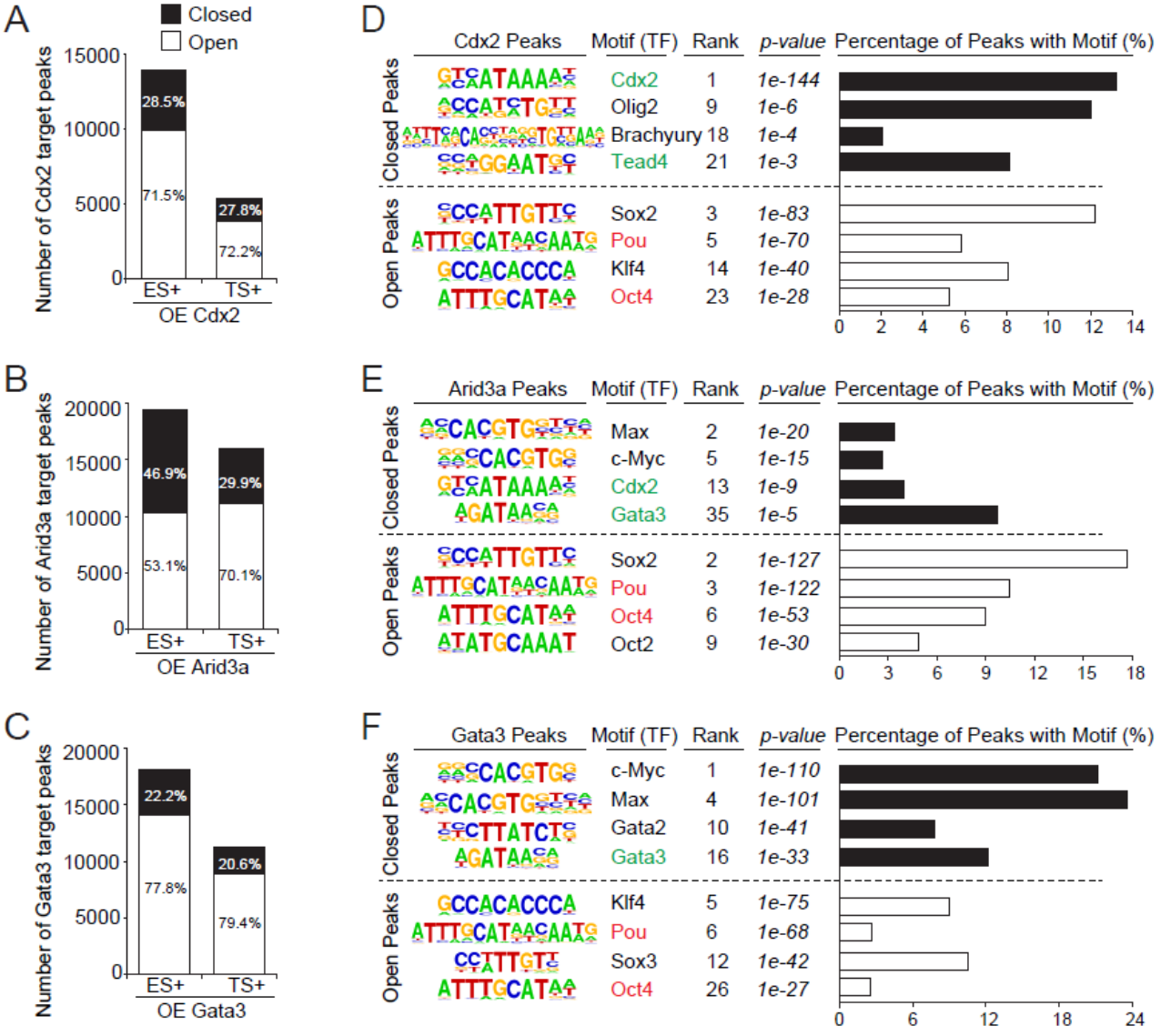


Figure 4.7. Enrichment of CAG binding motifs depends on chromatin landscape. (A-C) Bar graphs showing the number of Cdx2 (A), Arid3a (B), and Gata3 (C) target genes whose regulatory regions are associated with open or closed chromatin at the different stages of the reprogramming. (D-F) Upon OE of individual CAG factors in ES cells, significantly enriched motifs of Cdx2 (A), Arid3a (B), and Gata3 (C) binding sites that reside within open or closed chromatin are shown. Bar graphs (right panel) provide the percentages of individual CAG binding site with the motifs.

occupied by CAG factors (Figure 4.6A). Overall, our analysis revealed significant repression of ES cell-specific genes along with a slight induction of TS cell-specific genes at the early stage of programming; furthermore, the repression is mediated by direct binding of CAG factors (Figure 4.6B-E).

Each CAG factor binds both open and closed chromatin during the early stage of reprogramming regardless of the presence of canonical motifs

Given the importance of chromatin in global gene regulation and the fact that moderate changes in the chromatin landscape occurred shortly after CAG factor OE, we sought to determine whether CAG factors bind to open or closed chromatin. Previous studies introduced the "pioneer factor" concept as factors capable of binding directly to heterochromatin to activate silenced target cell-specific genes during reprogramming (Iwafuchi-Doi et al., 2016; Soufi et al., 2015). In order to investigate whether CAG factors function as pioneer factors and to what degree CAG factors bind to open and closed chromatin, we examined chromatin openness of individual binding sites during the reprogramming process. We observed that individual CAG factors prefer to bind open rather than closed regions of chromatin (Figure 4.7A-C); i.e., ~70% of Cdx2 and Gata3 binding events occurred within open regions (Figure 4.7A,C). This suggested that CAG factors are distinct from pioneer factors. Unexpectedly, however, all CAG factors show fewer peaks at the later stage of reprogramming (TS+) compared to the earlier stage (ES+), indicating that CAG factors might be crucial during the initiation of reprogramming but play a more limited role thereafter (Figure 4.7A-C). Next we examined the positions of CAG factor open vs. closed chromatin genomic binding sites relative to TSSs. We found that regardless of chromatin status, CAG factors preferentially bound to distal regulatory elements (Figure 4.8A,B).

Taken together, these results suggested that each CAG factor not only functions similar to pioneer factors by binding to closed chromatin for the activation of the desired (TS) cell type-specific genes, but also acts as a transcriptional repressor by binding to

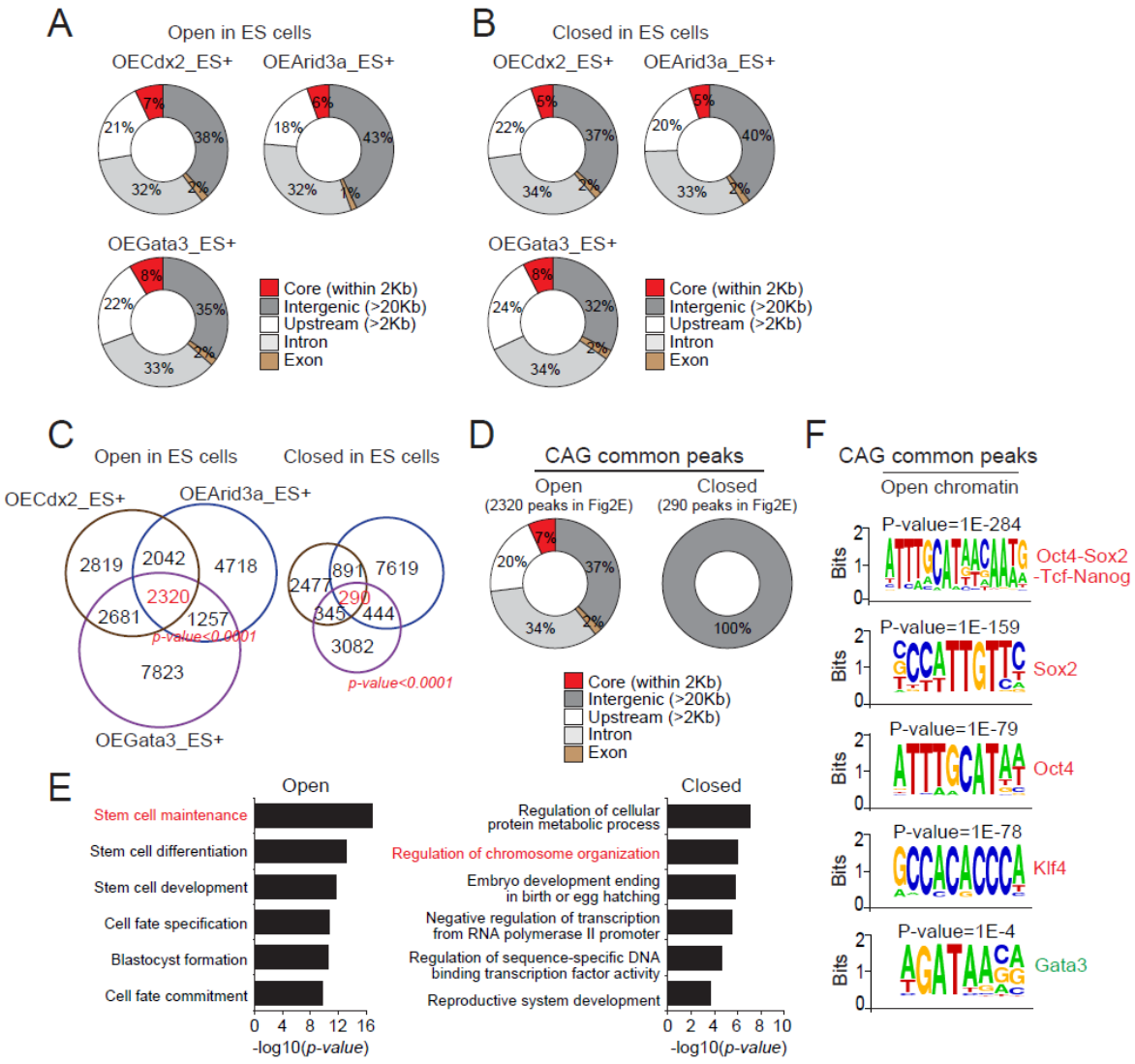


Figure 4.8. CAG factors share a significant number of target sites regardless of chromatin accessibility. (A,B) Pie charts presenting the distribution of each CAG factor binding sites that are open (A) or closed (B) in ES cells across the genome upon OE of individual CAG factor in ES cells. (C) Venn diagrams showing overlaps of open (left) or closed (right) loci among each CAG factor OE cells. (D) Pie charts showing the distribution of open (left) or closed (right) loci across the genome that are overlapped with the binding sites of CAG factors in ES cells. (E) Gene ontology (GO) analysis showing enriched biological process terms of common targets of CAG factors whose regulatory elements are either open (left) or closed (right). (F) De novo motif showing enriched motifs in common peaks of CAG factors.

open chromatin to suppress active ES cell-specific genes. To further explore this hypothesis, we examined whether each CAG factor bound its own unique targets or whether they all shared common binding sites. Comparison of individual CAG factor binding sites revealed that CAG factors share a significant number of target sites regardless of chromatin accessibility (Figure 4.8C). CAG common peaks within open chromatin are predominantly located at distal elements as expected (91%), whereas closed chromatin peaks are exclusively located within distal elements (100%) - a strong indication of pioneer factor-like activity (Figure 4.8D). Common open chromatin loci occupied by all CAG factors are primarily ES cell-specific, as evidenced by GO term enrichment (e.g., stem cell maintenance and cell fate specification) and DNA-binding similarity with ES cell-specific TFs (Figure 4.8E,F). This suggested that CAG factors promote reprogramming by occupying open regulatory regions of the ES cell-specific genes prior to repressing them, and that each individual factor shares a similar mechanism of repression in terms of target genes. On the other hand, GO analysis revealed that common closed chromatin-associated targets are associated with the regulation of chromosome organization and other chromatin regulatory-related terms (Figure 4.8E). Thus, CAG factors appear to regulate the chromatin landscape by activating expression of genes implicated in chromatin remodeling or modification.

To further address this interpretation, we performed motif analyses of individual CAG binding sites associated with the above observed open and closed chromatin regions. As shown in Figures 4.7D-F, top-ranked motifs associated with open CAG binding sites are ES cell-specific motifs (such as Oct4 and Sox2), whereas closed CAG target sites are enriched in TS cell-specific motifs (including Cdx2, Gata3 and Tead4). Unexpectedly, individual CAG binding sites were not enriched with chromosome organization-related motifs, indicating unlike their common targets, their unique targets might have high portion associated with TS cell-specific genes. Of particular note, CAG factors showed stronger occupancy signals at open rather than closed chromatin regions, even though these closed sites do not contain canonical CAG binding motifs (data not shown). These results suggest that CAG factors might aid binding of additional TFs by

occupying target sites indirectly via protein-protein interaction or competition with core pluripotency factors.

Arid3a promotes cell fate conversion by repression of ES cell-specific genes followed by activation of TS cell-specific genes

Among the CAG factors, Arid3a OE produced TS-like cells most similar to bona fide TS cells at the transcriptional and epigenetic level compared to OE of Cdx2 or Gata3 (Figure 4.2, 4.3). Therefore, we investigated Arid3a OE-mediated reprogramming in greater depth. We performed a *de novo* motif search for Arid3a binding sites over the course of reprogramming using the HOMER motif search tool (Heinz et al., 2010). Consistent with our previous study (Rhee et al., 2014), the motif constituting the predominant binding sites for Arid3a in ES cells is highly similar to the motif employed by Oct4 (Figure 4.9A). Unanticipatedly, binding motif preferences of Arid3a gradually changed over time during reprogramming (Figure 4.9A). This prompted another HOMER motif search (Heinz et al., 2010) for previously established TF motifs spanning the binding sites of Arid3a. The analysis revealed that motifs for ES cell-specific pluripotency factors are enriched during the earlier time point of reprogramming (OEArid3a_ES+)(Figure 4.9B). Conversely, TS cell-specific motifs were overrepresented at the later reprogramming time point (OEArid3a_TS+)(Figure 4.9B). These data clearly demonstrate that Arid3a dynamically switches its target site specificity during the course of reprogramming, perhaps by undergoing conformational changes, switching its protein interaction partners, or by recognizing degenerate motifs.

To further investigate this gradual shift in target site specificity and to elucidate its effect on repression, we assessed Arid3a occupancy over the course of reprogramming as well as the presence of H3K27ac - a histone mark enriched among active enhancers (Creyghton et al., 2010). We distinguished Arid3a target sites at different stages of reprogramming by clustering analysis. As shown in Figure 4.10A, we observed seven distinct patterns of Arid3a occupancy during the course of cell fate conversion. Class I and IV targets are mainly associated with stem cell development (Figure 4.10B). Strong

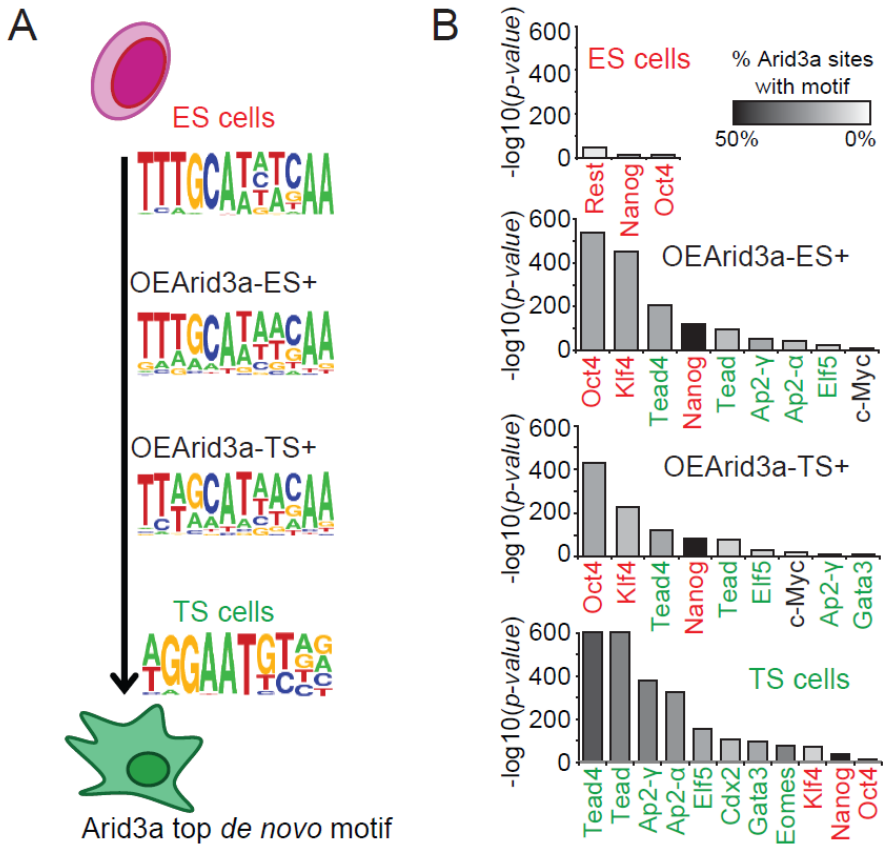


Figure 4.9. Arid3a dynamically switches its target sites specificity during the course of reprogramming. (A) *De novo* motif analysis of Arid3a binding sites showing the transitions of Arid3a motifs during the conversion of ES to TS-like cells. (B) Bar graphs of enriched motifs within Arid3a binding sites in ES, TS, and two different stages of reprogrammed ES cells.

Arid3a occupancy signals were predominantly observed in the regulatory regions of genes in these two classes at the early reprogramming time point (ES+). This was consistent with our observations that CAG factors initially mediate repression (Figure 4.10A). However, at the later time point (TS+), Arid3a occupancy of class I genes gradually disappears as occupancy shifts to class II and III genes (Figure 4.10A). In contrast to classes I and IV, genes belonging to class II and III are generally associated with TS cell-specific functions, such as placenta and labyrinthine development (Figure 4.10B). These data reveal dynamic changes in global Arid3a occupancy; i.e., Arid3a

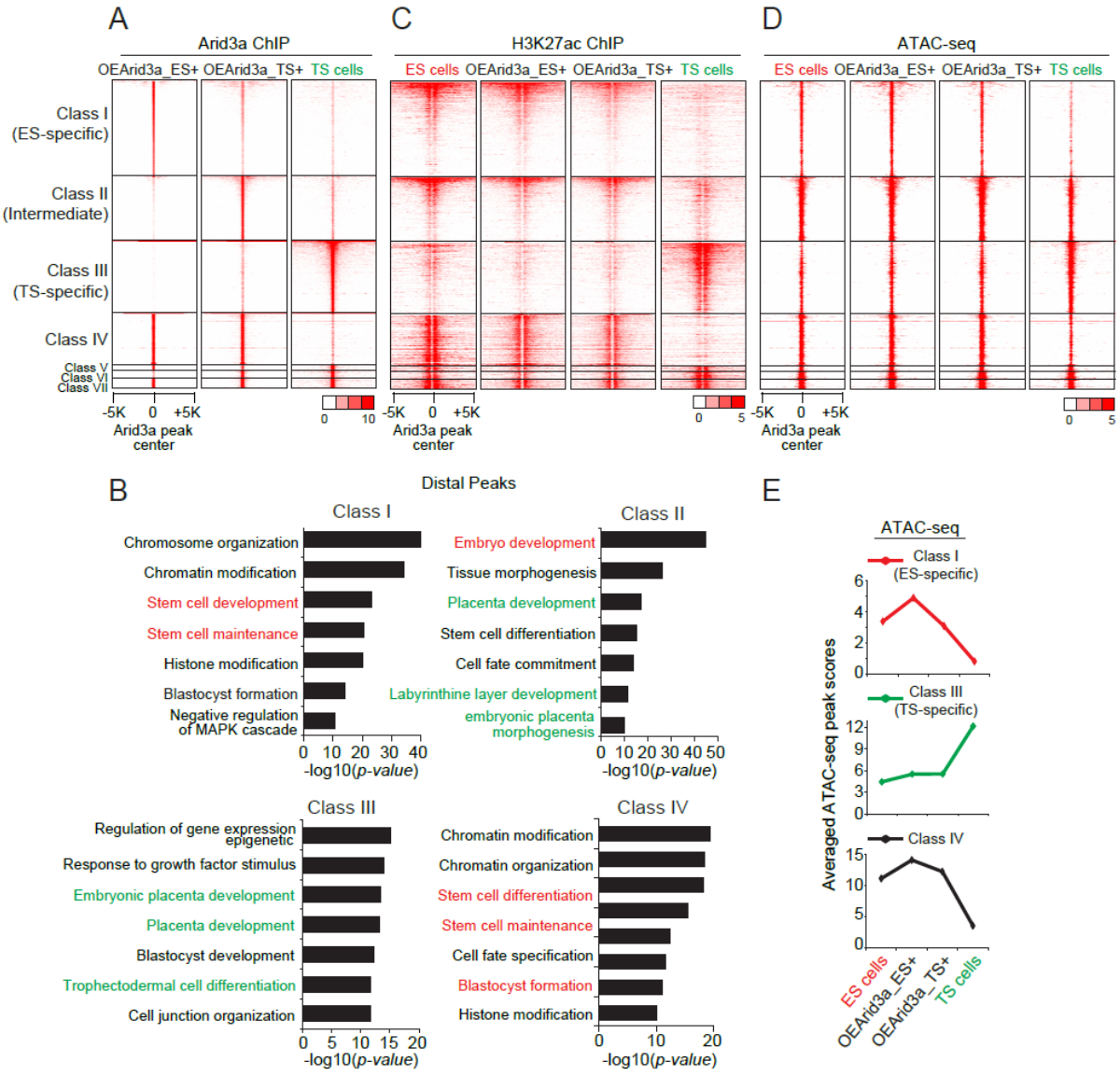


Figure 4.10. Arid3a promotes the reprogramming by repression of ES cell-specific genes followed by activation of TS cell-specific cells. (A,C,D) Heatmaps showing multiple clustered Arid3a binding sites (A), H3K27ac signatures (C), and chromatin landscapes that are dynamically changed depending on Arid3a occupancy during the reprogramming. (B) Significantly enriched terms of biological processes in the distal Arid3a binding sites of different classes by GO analysis. (E) Chromatin landscape changes in concert with changes in Arid3a occupancy plotted as a line graph showing average ATAC-seq peak scores within regions of Arid3a binding site classes shown in Figure 4.10A.

initially occupies ES cell-specific genes, presumably to repress them, followed by occupying TS cell-specific genes for their activation.

Since the H3K27ac signature is associated with active enhancers, we inspected its enrichment near Arid3a binding sites associated with open chromatin. We observed a strong positive correlation between the H3K27ac signature and Arid3a occupancy at each stage of reprogramming (Figure 4.10A,C). Class I-associated genes are ES cell-specific, and thus, are associated with H3K27ac marks. However, as reprogramming proceeded, these H3K27ac marks gradually disappeared from class I peaks (Figure 4.10C). Thus, we suggest that the initial binding of Arid3a to class I peaks ultimately removes H3K27ac signals to promote the suppression of ES cell-specific genes. Conversely, genes associated in class III gradually gain H3K27ac marks along with Arid3a occupancy during the course of reprogramming, indicating that Arid3a binding positively regulates these TS cell-specific genes (Figure 4.10C). Taken together, our results indicate that Arid3a represses ES cell-specific genes by reducing levels of H3K27ac, while it activates TS cell-specific genes by increasing H3K27ac. Since Arid3a itself does not possess acetyltransferase or deacetylase activity, our results imply that Arid3a may recruit proteins with these enzymatic activities. The feasibility of such a mechanism was provided by our demonstration (Rhee et al., 2014) that Arid3a recruits Hdac1 and Hdac2 to catalyze target repression.

Consistent with changes in Arid3a occupancy during reprogramming, we found that chromatin accessibility is also altered (Figure 4.10D,E). Average ATAC-seq peak scores revealed that genes in classes I (ES cell-specific) and IV showed decreased accessibility along with loss of Arid3a occupancy during reprogramming, implying a gradual conversion to heterochromatin. Alternatively, class III (TS cell-specific) genes gained chromatin accessibility as Arid3a occupancy increased, consistent with conversion to euchromatin—consistent with classical pioneer factor activity (Figure 4.10E). Thus, during the course of reprogramming, Arid3a switches its distal regulatory targets from ES

cell-specific to TS cell-specific genes, resulting in a global shift in the transcriptional and epigenetic landscape from an ES to a more TS-like profile.

CAG factors deactivates pre-existing ES cell-specific enhancers

Based on our finding that individual CAG factors preferentially bind to loci distal from the TSSs of their target genes and that ES cell-specific genes are directly repressed by CAG factors, we reasoned that CAG factors might regulate the activity of ES cell-specific enhancers. To investigate this possibility, we analyzed the overlap of CAG factor targets with ES cell-specific enhancers, as defined by the histone enhancer mark, H3K27ac, as well as by ES cell-specific "super-enhancers"—a cluster of enhancers densely occupied by the ES cell-specific pluripotency factors, P300, and Mediator (Whyte et al., 2013). We observed that >70% of CAG factor targets overlap with active enhancers defined by H3K27ac signals and that ~50% of ES cell-specific super-enhancers are also occupied by CAG factors upon their OE in ES cells (Figure 4.11A,B). This indicated that CAG factors prefer to occupy active ES cell-specific enhancers when expressed in ES cells. In addition, binding sites for the ES cell-specific TF Oct4, previously defined in ES cells, also displayed significant overlap with CAG binding sites (Figure 4.11C,D).

Based on these observations, we addressed two potential mechanisms of CAG-mediated repression of ES cell-specific genes. First, CAG factors might compete with ES cell-specific TFs, such as Oct4, to bind to ES cell-specific enhancers. To test this hypothesis, we determined Oct4 occupancy upon OE of CAG factors in ES cells. Indeed, Oct4 occupancy was dramatically decreased, suggesting that there may be some level of binding competition among Oct4 and CAG factors at ES cell-specific enhancers (Figure 4.11E). However, we also observed decreased levels of Oct4 expression during the course of reprogramming (data not shown); thus, we cannot rule out the possibility that this decreased Oct4 occupancy is a result of reduced levels of Oct4.

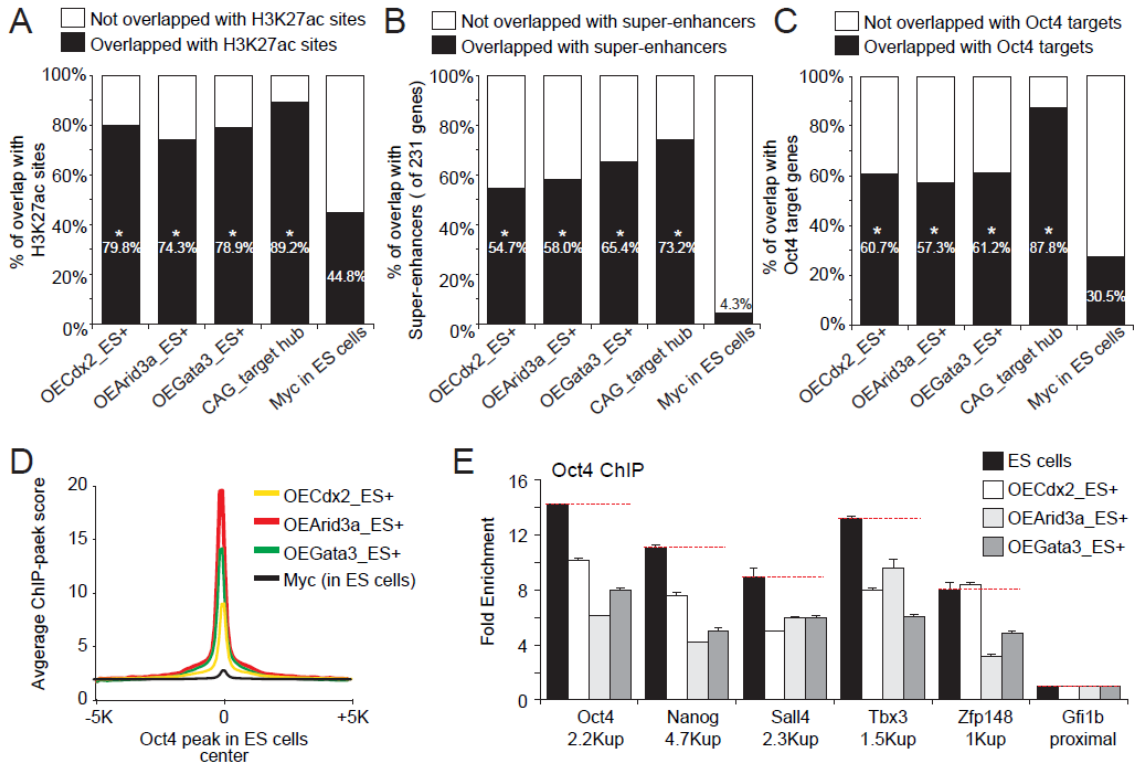


Figure 4.11. CAG factors decommission ES cell-specific enhancers. (A-C) Bar graphs depicting overlaps of individual CAG factor binding sites and target hub upon OE of each factor in ES cells with H3K27ac enriched sites (A), super-enhancers (B) and Oct4 binding sites (C) in ES cells. Myc binding sites in ES cells are used as negative control. Asterisks mark significant overlaps (*P*-value < 0.0001). (D) Average occupancy profiles of each CAG factor centered on Oct4 binding sites in ES cells following OE of CAG factors in ES cells. (E) Oct4 ChIP-qPCR plotted to show relative Oct4 occupancy in the regulatory regions of ES cell-specific genes upon OE of CAG factors. Error bars depict standard deviations of biological triplicates.

The alternative interpretation is that CAG factors deactivate ES cell-specific enhancers by modifying histone signatures. To test this possibility, we overexpressed Arid3a in ES cells using a doxycycline (dox) - inducible system. Cells were treated with dox for 4 days, followed by withdrawal and further culturing for 4 additional days. Dox-dependent induction of Arid3a was confirmed by the expression of Zsgreen1 (Figure 4.12A). Next, we determined by quantitative PCR whether enhancer-associated H3K27ac

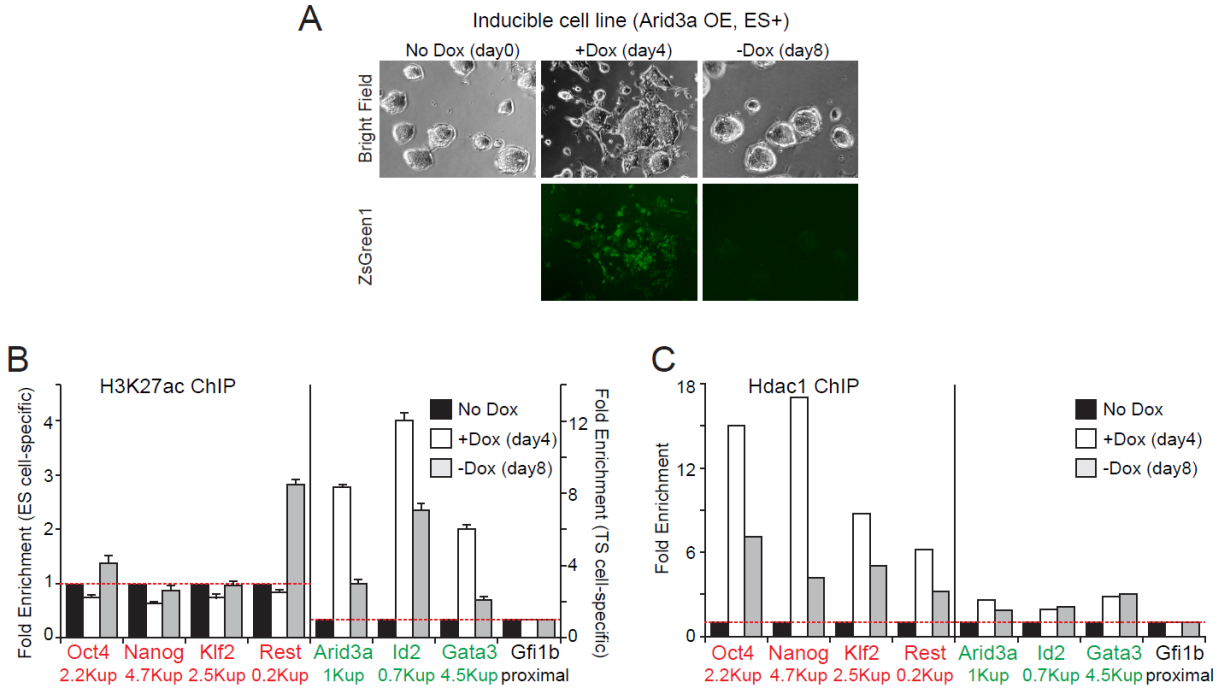


Figure 4.12. OE of Arid3a in ES cells initiates ES to TS-like cell reprogramming through changes in histone signatures on the enhancers. (A) Cell morphologies of ES cells and Arid3a-inducible ES cells upon either treatment with doxycycline (dox) for 4 days or following withdrawal of dox. (B,C) H3K27ac (B) and Hdac1 (C) ChIP-qPCR showing relative levels of H3K27ac in the regulatory regions of ES cell-specific genes (red) and TS cell-specific genes (green).

signatures were modulated by CAG factors. As shown in Figure 4.12B, upon induction and removal of dox, we observed modest changes in H3K27ac marks within ES cell-specific enhancers. In contrast, induction of Arid3a significantly increased the levels of H3K27ac at the enhancers of TS cell-specific genes. Finally, upon removal of dox, levels of H3K27ac decreased, indicating Arid3a-dependent regulation of H3K27ac on TS cell-specific genes.

Next, we previously determined that Arid3a interacts with Hdac1. Since HDACs are responsible for deacetylation of H3K27ac (Urvalek and Gudas, 2014), we performed Hdac1 ChIP in these Arid3a OE cells shown in Figure 4.12A, followed by qPCR to monitor the enrichment of Hdac1 at regulatory regions of ES cell-specific as well as at TS cell-specific genes in these cells. We observed strong enrichment of Hdac1 at the

enhancers of ES cell-specific genes but no enrichment of Hdac1 at the regulatory elements of TS cell-specific genes in early stage of the reprogramming (Figure 4.12C). Conversely, the shutdown of Arid3a induction by dox withdrawal reduced enrichment of Hdac1 at ES cell-specific enhancers (Figure 4.12C). Together, our results provide evidence that OE of Arid3a in ES cells initiates ES to TS-like cell reprogramming through changes in histone signatures on the enhancers.

4.4 DISCUSSION

Reprogramming is lineage-specific, TF-mediated conversion of one cell type to another. Yet the underlying mechanism by which TFs initiate and complete the reprogramming process has remained elusive. We show here that ectopic expression of CAG factors in ES cells promotes dynamic changes in global gene expression as well as within the chromatin landscape during the early stages of reprogramming to TS-like cells. Previous studies proposed the concept of pioneer factors, which initially bind preferentially to heterochromatin of cell-type-specific genes and further remodel this "closed" chromatin structure into an accessible, "open" version. CAG factors occupy closed chromatin within ES cells to activate silenced genes essential for TS cell identity, suggesting that they partially have pioneer factor-like activity. However, surprisingly, CAG factors also repress ES cell-specific genes. Upon OE, they prefer to bind within open rather than closed chromatin of ES cells to repress genes essential for maintenance of pluripotency. Consistent with many TFs, CAG factors bind to distal regulatory elements (i.e., enhancers) and modulate epigenetic features such as histone modifications to shift the entire transcriptional program to that of the target cell type.

Consistent with previously identified pioneer factors which bind to DNA via recognition of specific sequence motifs, our analyses revealed that the binding sites of CAG factors within closed chromatin contain degenerate, "partial" versions of their canonical motifs. Conversely, their binding sites within open chromatin lack canonical CAG motifs, suggesting that CAC factors may require interactions with other factors to facilitate their recruitment to these non-consensus open target sites. Crucial to this

argument is our observation that CAG factor binding appears to convert open regions of chromatin within ES cells to heterochromatin within TS-like cells. This leads to repression of ES cell-specific genes as well as accessibility of closed regions of chromatin corresponding to genes involved in TE development. Our integrative analysis of CAG factor occupancy, gene expression, and chromatin accessibility at different stages of early reprogramming revealed a timely and ordered binding strategy, followed by alterations within chromatin landscapes as well as dramatic changes in global gene expression.

These observations led us to a step-wise model in which, initially, CAG factors predominantly occupy open chromatin of ES cell-specific genes to deactivate them. Then, they switch their binding site preference onto the regulatory elements of TS cell-specific genes for activation. Both deactivation and activation appear to be associated, if not mediated, by histone modifications, as levels of H3K27ac within enhancers change significantly over the course of reprogramming. This mechanism requires dynamic and sequential occupancy changes within the binding of CAG factors, as they must adjust their specificity from ES cell-specific to TS cell-specific in a timely, ordered manner. To our knowledge, this is the first example of a step-wise mechanism of TF-mediated reprogramming elucidated at both transcriptional and epigenetic levels. These events, which underlie the reprogramming of ES cells to TS-like cells, uniquely expand the established mechanisms of pioneer factor-mediated transactivation. However, this is yet limited to reprogramming from ES cells to TS-like cells.

Recent studies have employed mouse embryonic fibroblasts (MEFs) to establish so-called “induced” trophoblast stem cells (iTS) cells (Benchetrit et al., 2015; Kubaczka et al., 2015). However, there are unique advantages in employing ES cells to generate functional TS-like cells. Compared to MEFs, mouse ES cells contain far more open chromatin structure (Mikkelsen et al., 2007). Indeed, both ES cell-specific genes as well as some TS cell-specific genes are accessible in ES cells. Such configuration might facilitate not only the binding of CAG factors, but the time required for reprogramming.

For example, generation of iTS cells from MEFs usually takes ~3 weeks, whereas reprogramming of ES to TS-like cells takes less than one week to initiate changes in transcriptional programs and to generate functional TS-like cells *ex vivo* (Rhee et al., 2014). In addition, this sequential step-wise mechanism may be due to the fact that closed chromatin - the status of repressed TS cell-specific genes in MEFs - delays reprogramming.

Taken together, our systematic approach of analyzing CAG occupancy, transcriptomic and epigenetic changes during reprogramming of ES to TS-like cells broadens our understanding of the mechanism underlying TF-mediated reprogramming. Our stepwise and integrative dissection of the initial steps of reprogramming may help improve both the efficiency of the reprogramming process as well as facilitate stem cell therapies in patients whom rely on changing the fate of their own cells for treatment.

CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS

SUMMARY

During my doctoral studies, I have investigated transcriptional and epigenetic regulation of the first cell fate decision during early embryonic development and ES to TS-like cell reprogramming. I have employed a broad panel of molecular biological approaches, along with high-throughput genomics and computational analyses. Together, I have attempted to present a holistic picture by addressing mechanisms of development and reprogramming and their validation *in vivo* and *ex vivo*.

First, I observed that Arid3a, a previously characterized transcriptional activator of B-cell development, is a critical regulator for controlling the first cell fate decision and placental development. I demonstrated that ectopically expressed Arid3a drives induction of TE-like transcriptional programs in ES cells. Arid3a is not only required for the maintenance of self-renewal of TS cells, but also promotes further trophoblastic lineage differentiation. I determined that, consistent with the *in vitro* results, *Arid3a*$^{-/-}$ mice suffer severely impaired post-implantation development, which results in early embryonic lethality. Mechanisms underlying this essential Arid3a function include regulated nuclear entry-export, HDAC-mediated deacetylation of target genes, and regulation of embryonic hematopoiesis. Collectively, my studies revealed that Arid3a functions as a pivotal regulator of TE and placental development by regulating commitment to the first cell fate as well as by the execution of TE lineage differentiation.

Second, I studied the mechanisms underlying transcription factor (TF)-mediated trans-differentiation/direct reprogramming of mouse ES to TS-like cells. I addressed this mechanism specifically via the TS cell-specific TFs Cdx2, Arid3a, and Gata3 (CAG factors). I carried out time–course profiling of chromatin accessibility, transcriptomes, and occupancy of reprogramming factors during direct ES to TS-like cell conversion. I discovered that CAG factors orchestrate conversion via a two-step mechanism: 1) initial repression of ES cell-specific genes by decommissioning of active enhancers and 2) direct activation of TS cell-specific genes via accession of closed chromatin by

employing different DNA binding motifs during reprogramming. My data support a sequential gene regulation mechanism in which the exit of ES cells from pluripotency states is followed by the activation of TS cell-specific genes, leading to the conversion of ES cells to multipotent TS-like cells.

UNDERSTANDING EPIGENETIC MARKS AND SIGNALING PATHWAYS OF THE ES TO TS-LIKE CELL REPROGRAMMING

Although it has been established that several TFs (~9 TFs) play critical roles in the first cell fate decision, it is unknown when and how epigenetic marks arise and fade away during this crucial point in development. It is also unknown what role such marks play during ES to TS-like cell reprogramming. This is vital to consider, as histone modifications have vast implications for gene expression by promoter silencing, RNA polymerase recruitment, and other mechanisms. This question can be addressed by identifying essential epigenetic modifiers during the ES to TS-like cell transition, including their target sites within the genome and determination of the mechanism(s) by which they are recruited or suppressed by cell type-specific TFs. Understanding such developmentally regulated epigenetic signatures will be crucial in unraveling key features of the previously observed epigenetic plasticity during this transition. Progress in this area is essential for the ultimate goal of developing safe cell therapies and regenerative medicines.

In addition to discerning ES to TS-like cell reprogramming of transcriptional and epigenetic regulation, understanding the signaling pathways governing self-renewal and multipotency in mouse TS cells is crucial. Although the Ras/Mapk and Nodal signaling pathways are strongly implicated, each signaling pathway participates in crosstalk with other pathways; thus, it is vital to establish a signaling network to better understand how distinct trophoblast lineages are specified—e.g., what TFs regulate differentiation to giant cells vs. labyrinthine cells. The benefits of constructing such a network are numerous. After all, the use of chemical inhibitors that target specific signaling pathways has allowed the derivation of pluripotent ES cells across species and facilitated creative

approaches to cell reprogramming. Additionally, understanding signaling in the ES to TS-like cell reprogramming context may enable development of a chemically defined media to culture and maintain TS cells. This would eliminate the variability and ethical issues inherent in serum use, as well as link individual signaling components to vital transcriptional targets.

GENERATION OF HUMAN TS CELLS

Although derivation of human TS cells has failed to date, human TS cells can provide a highly valuable tool to study genetic and epigenetic control mechanisms of trophoblast proliferation, differentiation, and function in normal and pathological placentation during human embryogenesis. Human TS cells may also provide a huge benefit for human health. The failure of viable derivation of human TS cells highlights the critical need for understanding embryonic development on a molecular level. Nevertheless, and similar to what we observed in mouse ES cells, transient induction of ARID3A can up-regulate TE-associated factors such as GATA3 and TFAP2C while down-regulating levels of OCT4 and NANOG. Thus, the regulatory function of mouse Arid3a during placental development may be conserved between mouse and human. This means that our discoveries regarding the function of Arid3a in mouse reprogramming and hematopoiesis would provide helpful starting points for examining their role in humans. In addition, in the future, these studies might be lead to isolation and culturing of human TS cells.

Unfortunately, the human implantation process is still poorly understood. The loss of potential embryos occurs relatively frequently due to the loss of blastocysts before or around implantation. Thus, it will be essential to decipher how TS cell-specific TFs are conserved in the human placentation process and to examine how their deregulation affects serious pregnancy diseases, such as pre-eclampsia. Once viable human TS cells are generated, they may be used in the future to treat the many pregnancy-related disorders that appear to involve dysregulated trophoblast differentiation.

TS CELL TRANSCRIPTIONAL NETWORK

Proper trophoblast differentiation is essential for placental development and successful mammalian reproduction. Paradoxically, despite its fundamental importance, the placenta remains one of the least understood organs in mammals. There are only a few TFs identified to date that are implicated in trophoblast lineage specification. This short list includes ARID3A. Although the ES cell pluripotency network is well-characterized, TS cell regulatory networks remain insufficiently understood. Also, it is unknown how TS cell-specific TFs regulate one another to maintain self-renewal and multipotency. We hypothesize that, analogous to ES cells, TS cells are maintained via positively regulated feedback loops governed by multiple TS cell-specific master TFs. Perturbations in this network then cause an exit from multipotency and result in terminal differentiation into specialized placental cells. This is crucial to consider, as improper or unbalanced differentiation of placental lineages is a hallmark in placental disorders. Indeed, such differentiation defects also pose risks for congenital malformations, pre-eclampsia, and miscarriage.

In the age of genome sequencing and massive biological data sets, systems-level approaches have become increasingly critical for understanding biological processes. Construction of a TS cell-specific regulatory network would be a great asset to the fields of development, differentiation, and medicine. It is our great hope that the present studies delineated in this dissertation may catalyze further advances in placental biology and trophoblast development.

APPENDIX

APPENDIX A. RT-qPCR PRIMERS

	Forward	Reverse
Arid3a	AGGTTATCAACAAGAAACTGTGGAG	TACTTCATGTACTGTGTCCGAAGTG
Bmp2	CGCTTCTTCTTCAATTTAAGTTCTG	AACTACTGTTTCCCAAAGCTTCCT
Bmp5	AGCCTACATGATACCAACTTTCTGA	GATCAAACCGAAACTCTTTGTAATG
Cdx2	GCGAAACCTGTGCGAGTGGATG	CGGTATTTGTCTTTTGTCCTGGTTTTCA
Cxcr4	ATGGAACCGATCAGTGTGAGTAT	AAGATCCTATTGAAATGGACGTTTT
Dlx3	CTCACACAAACACAGGTGAAAATCT	ATGGAGTCACTGTTGTTGGGG
Fgf5	GGATTGTAGGAATACGAGGAGTTTT	AACTTACAGTCATCCGTAAATTTGG
Gapdh	AAATTCAACGGCACAGTCAAG	CACCCCATTTGATGTTAGTGG
Gata2	GCCTCTACCACAAGATGAATGG	GTCTGACAATTTGCACAACAGG
Gata3	TGGGCTGTACTACAAGCTTCATAA	CTTTTTCGATTTGCTAGACATCTTC
Gata4	TTCTCAGAAGGCAGAGAGTGTGT	ATGCCGTTCATCTTGTGATAGAG
Gata6	GACGGCACCGGTCATTACC	ACAGTTGGCACAGGACAGTCC
Gcm1	TGTTGAGCAGACCTTATCATGGAA	GTCAGTCGTTTTCACGTTCTGAG
Gsc	AGAAGGTGGAGGTCTGGTTTAAG	GAGGACGTCTTGTTCCACTTCT
Hand1	CCTTCAAGGCTGAACTCAAAAA	GCGCCCTTTAATCCTCTTCT
Hoxb1	AGGTCAAGAGAAACCCACCTAA	AAATGAAATTCCTTCTCCAGCTC
Id2	ATCACCAGAGACCTGGACAGAAC	GCTATCATTCGACATAAGCTCAGA
Isl1	GGGATGGGAAAACCTACTGTAAAAGAGA	GTCGTTCTTGCTGAAGCCTATGCTG
Klf4	CTGCAGCTTGCAGCAGTAAC	AGCGAGTTGGAAAGGATAAAGTC
Krt8	AGAACATGAGCATTCATACGAAGA	GAGCTCATTCCGTAGCTGAAG
Nanog	AGGGTCTGCTACTGAGATGCTCTG	CAACCACTGGTTTTTCTGCCACCG
Nes	AGGACCAGGTGCTTGAGAGA	TTCGAGAGATTCGAGGGAGA
Oct4	CTGAGGGCCAGGCAGGAGCACGAG	CTGTAGGGAGGGCTTCGGGCACTT
Otx2	AAGTGAGTTCAGAGAGTGGAACAAG	CTCCAGATAGACACTGGAGCACT
Sox17	CTAAGCAAGATGCTAGGCAAGTCT	GTACTTGTAGTTGGGGTGGTCCT
Sox2	GCGGAGTGGAAACTTTTGTC	TATTTATAATCCGGGTGCTCCT
T	CTTCAAGGAGCTAACTAACGAGATG	GTCCAGCAAGAAAGAGTACATGG
Tcfap2c	CCACGTCGAAGTACAAAGTAACTG	TTTTTGGACTTTGCTCTTCTGAG
Tpbpa	GCAAGAGCAGAAGGATAAAGAAGTT	CACTCATTTCTATGTTGGAGCCTTC
Wnt2	TTAATATGAACGTCCCTCTCGGTG	GTAGCTCTCATGTACCACCATGAA
ARID3A	GGACTCTGCTGTGAACCTGAC	GGCTGAGCAAACAGAACTCCT
CDX2	CTGGAGAAGGAGTTTCACTACAGTC	AACCAGATTTTAACCTGCCTCTC
GATA3	GGGCTCTACTACAAGCTTCACAATA	CACTTTTTGGATTTGCTAGACATTT
GCM1	GACGCTTTATATTTTTCCAGTCAAA	AAGGTGCTGTGTTCACTTTCTTC
OCT4	AAGGAGAAGCTGGAGCAAAAC	GTATATCCCAGGGTGATCCTCTT
NANOG	CCAGAACCAGAGAATGAAATCTAAG	AAGAGTAAAGGCTGGGGTAGGTAG
GAPDH	TTGGAACATGTAAACCATGTAGTTG	AGTCAACGGATTTGGTCGTATT

APPENDIX B. CHIP-QPCR PRIMERS

	Forward	Reverse
Arid3a_3.8K_Down	ACTCTCATTCAGAGGCAGCAG	CCTGTCCTCATCTGTCATGGT
Arid3a_Proximal	CAAGATCCCCACAACAGTCTAAC	AAGTTGAGACGATTCTAGTGGGTAA
Cdh3_0.5K_Down	AAACTTAAAAGTGGAGGGTCATTTC	AAAATCCCACTGTCCTCTGTAGTC
Cdh3_3K_Up	GAAGAGCTGAGTATCGTTTTCTCC	CTAATTAAGTAAACACTCGGCTGGA
Cdx2_0.5K_Down	CCTCCTTGCCAGTAAGCTGTAG	GTCTTGTAAACACTCGTTAATCACGTA
Cdx2_2.2K_Down	TGATGTTAATGTAGGGAAGCTGTG	GGGCTGAAAGGTGTACATCAA
Cdx2_Proximal	GTCTTGTAAACACTCGTTAATCACGTA	CCTCCTTGCCAGTAAGCTGTAG
Fgf4_10K_Up	TCTCTTCCCAGAGGTGAAGGT	CAGGGTATATCTCTGTTCAGTCTCC
Gata3_4.5K_Up	GAGATGCGTTTAATTCGATTCC	TTCTTTTACAACTTCCATGCCTAAG
Gfi1b_Proximal	CGCCAGATTTTGACACAAATAA	CTGCACAGACAGACACTTCTCC
Id2_0.7K_Up	CTCAGTGACACGAATTGTAGGAAC	GAGTCATTATAGTAACCCTGCCTCA
Id2_Proximal	GCATTTACCAAACCAAACATTTAAC	AGCTTATGTCGAATGATAGCAAAGT
Klf4_1.9K_Down	GAGAATCTGCGCCACCTC	CGGAATGTATACTGGGTCCAAC
Klf4_1K_Down	CCTAATATTGACTTGTTGGGTGAAG	TCTTACCTTGGTGTTAGCTGAGATT
Nanog_0.2K_Up	GATGAATAAAGTGAAATGAGGTAAAGC	GTAATGCAAAAGAAGCTGTAAGGTG
Nanog_4.7K_Up	AAAGACACTAAAGAGGCAGGACAG	GTTTACCCCAAGTTCTACAAAGGTT
Nanog_4K_up	CATTTCTTTCCTTAGTGAGATTGGA	ATATTGAAGGACAATAGGCATTCTG
Oct4_2K_Up	CATAACAAAGGTGCATGATAGCTC	ACAAAGCTTCCTCAATAGCAGATTA
Oct4_Proximal	GTCTGGAAGACACAGGCAGATAG	ACCCTCTAGCCTTGACCTCTG
Pecam1_9K_Up	GCTGCAATCAATTTTATTTAACCAG	TAATGCTAAATTCTGTACCCAGAGG
Rest_1K_Down	TTTATTCTTTGCATATCTTGCCTTT	TGAGTTTCTAAATGGTCTCAACCTC
Rest_3.4K_Up	TGATTTTCTCCTGGTTAAGAGTTTG	AACAAAATCTAAGTGCCTGCCTAC
Sall4_2.3K_Up	CTGCTCAAGGCATTGTAAGCTAT	AGACAGAAGTGGAAGAAGCCTTT
Tbx3_0.4K_Up	GCAGGAGCTAGAGGATCTGACT	GCCAATAAGCCTTAACAAAAGC
Tbx3_0.5K_Down	CATGAGACTATGCACCTTGTATTTG	GGTGGTTTTAAACACTAGTCATTGG
Tcfap2c_4.7K_Up	GTCATAAAATCTCTTTGTGCCCTTA	GTATAAAAGGAAACACAGACAAAGAGG
Zfp42_Proximal	GCATCCTCTGCTTGTGTAAATTC	CTCAGTTATGCAAATGCCTCTTC

APPENDIX C. CLONING PRIMERS

	Forward	Reverse
FB_Arid3a	CGCGGATCCTCAAGCTGCAGGCTGTGATGGAG	CGCGGATCCTTAAGGCAAGGAGTTGTTTG
FB_Cdx2	CGCGGATCCTCTACGTGAGCTACCTTCTGG	CGCGGATCCTCACTGGGTGACAGTGGAGTTTAAAAC
FB_Gata3	CCGAGCTCGAGGATCCTCATGGAGGTGACTGCGGACCA	TAGAACTAGTGGATCCCTAACCCATGGCGGTGACCA
pEF1alpha_IRES_ZsGreen_Arid3a	CGCTGCTAGCATGAAGCTGCAGGCTGTGATGG	CGCTGAATTCTTAAGGCAAGGAGTTGTTTGAG
pEF1alpha_IRES_ZsGreen_Cdx2	CGCTGCTAGCATGTACGTGAGCTACCTTCTGG	CGCTGAATTCTCACTGGGTGACAGTGGAG
pMSCV_Arid3a	CTCTCTCGAGGCGATGAAGCTGCAGGCTGTGATGGAG	CTCTGAATTCGCGTTAAGGCAAGGAGTTGTTTGAGGTAG
pGEM-in situ-Arid3a	GCGGACCCCAAGAGGAAAGAGTT	CTGGGTGAGTAGGCAAAGAGTGAGC
Inducible_Arid3a	GCCCCCGGGACGCGTATGAAGCTGCAGGCTGTGATGGAG	CTACCCGGTAGAATTCTTAAGGCAAGGAGTTGTTTGAGGT

REFERENCES

Ai, D., Fu, X., Wang, J., Lu, M.F., Chen, L., Baldini, A., Klein, W.H., and Martin, J.F. (2007). Canonical Wnt signaling functions in second heart field to promote right ventricular growth. Proceedings of the National Academy of Sciences of the United States of America *104*, 9319-9324.

Alder, O., Lavial, F., Helness, A., Brookes, E., Pinho, S., Chandrashekran, A., Arnaud, P., Pombo, A., O'Neill, L., and Azuara, V. (2010). Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. Development *137*, 2483-2492.

Ananth, C.V., Keyes, K.M., and Wapner, R.J. (2013). Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. Bmj *347*, f6564.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B.*, et al.* (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183-197.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science *321*, 699-702.

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nature reviews Molecular cell biology *10*, 91-103.

Artus, J., and Chazaud, C. (2014). A close look at the mammalian blastocyst: epiblast and primitive endoderm formation. Cellular and molecular life sciences : CMLS *71*, 3327-3338.

Auman, H.J., Nottoli, T., Lakiza, O., Winger, Q., Donaldson, S., and Williams, T. (2002). Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. Development *129*, 2733-2747.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development *17*, 126-140.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research *37*, W202-208.

Bainbridge, S.A., Minhas, A., Whiteley, K.J., Qu, D., Sled, J.G., Kingdom, J.C., and Adamson, S.L. (2012). Effects of reduced Gcm1 expression on trophoblast morphology, fetoplacental vascularity, and pregnancy outcomes in mice. Hypertension *59*, 732-739.

Baussano, I., Tardivo, I., Bellezza-Fontana, R., Forneris, M.P., Lezo, A., Anfossi, L., Castello, M., Aleksandar, V., and Bignamini, E. (2006). Neonatal screening for cystic fibrosis does not affect time to first infection with Pseudomonas aeruginosa. Pediatrics *118*, 888-895.

Beck, F., Erler, T., Russell, A., and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. Developmental dynamics : an official publication of the American Association of Anatomists *204*, 219-227.

Benchetrit, H., Herman, S., van Wietmarschen, N., Wu, T., Makedonski, K., Maoz, N., Yom Tov, N., Stave, D., Lasry, R., Zayat, V.*, et al.* (2015). Extensive Nuclear Reprogramming Underlies Lineage Conversion into Functional Trophoblast Stem-like Cells. Cell stem cell *17*, 543-556.

Bernardo, A.S., Faial, T., Gardner, L., Niakan, K.K., Ortmann, D., Senner, C.E., Callery, E.M., Trotter, M.W., Hemberger, M., Smith, J.C.*, et al.* (2011). BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. Cell stem cell *9*, 144-155.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.
Bilban, M., Haslinger, P., Prast, J., Klinglmuller, F., Woelfel, T., Haider, S., Sachs, A., Otterbein, L.E., Desoye, G., Hiden, U.*, et al.* (2009). Identification of novel trophoblast invasion-related genes: heme oxygenase-1 controls motility via peroxisome proliferator-activated receptor gamma. Endocrinology *150*, 1000-1013.
Bischoff, M., Parfitt, D.E., and Zernicka-Goetz, M. (2008). Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. Development *135*, 953-962.
Blelloch, R., Wang, Z., Meissner, A., Pollard, S., Smith, A., and Jaenisch, R. (2006). Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. Stem cells *24*, 2007-2013.
Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.
Cahan, P., and Daley, G.Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. Nature reviews Molecular cell biology *14*, 357-368.
Capo-Chichi, C.D., Rula, M.E., Smedberg, J.L., Vanderveer, L., Parmacek, M.S., Morrisey, E.E., Godwin, A.K., and Xu, X.X. (2005). Perception of differentiation cues by GATA factors in primitive endoderm lineage determination of mouse embryonic stem cells. Developmental biology *286*, 574-586.
Carlson, L.L., Page, A.W., and Bestor, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes & development *6*, 2536-2541.
Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem cells *26*, 1496-1505.
Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.
Chawengsaksophak, K., James, R., Hammond, V.E., Kontgen, F., and Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. Nature *386*, 84-87.
Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Developmental cell *10*, 615-624.
Chen, C.P., Chen, C.Y., Yang, Y.C., Su, T.H., and Chen, H. (2004). Decreased placental GCM1 (glial cells missing) gene expression in pre-eclampsia. Placenta *25*, 413-421.
Chen, T., and Dent, S.Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature reviews Genetics *15*, 93-106.
Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.
Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nature genetics *45*, 325-329.
Cimmino, L., Abdel-Wahab, O., Levine, R.L., and Aifantis, I. (2011). TET family proteins and their role in stem cell differentiation and transformation. Cell stem cell *9*, 193-204.

Cormier, S., Vandormael-Pournin, S., Babinet, C., and Cohen-Tannoudji, M. (2004). Developmental expression of the Notch signaling pathway genes during mouse preimplantation development. Gene expression patterns : GEP *4*, 713-717.
Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.
Cross, J.C. (2005). How to make a placenta: mechanisms of trophoblast cell differentiation in mice--a review. Placenta *26 Suppl A*, S3-9.
Cross, J.C., Werb, Z., and Fisher, S.J. (1994). Implantation and the placenta: key pieces of the development puzzle. Science *266*, 1508-1518.
Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.
Dietrich, J.E., and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. Development *134*, 4219-4231.
Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H.*, et al.* (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. Cell stem cell *4*, 403-415.
Donnison, M., Beaton, A., Davey, H.W., Broadhurst, R., L'Huillier, P., and Pfeffer, P.L. (2005). Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. Development *132*, 2299-2308.
Dovey, O.M., Foster, C.T., and Cowley, S.M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. Proceedings of the National Academy of Sciences of the United States of America *107*, 8242-8247.
Dyce, J., George, M., Goodall, H., and Fleming, T.P. (1987). Do trophectoderm and inner cell mass cells in the mouse blastocyst maintain discrete lineages? Development *100*, 685-698.
Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.
Ezashi, T., Telugu, B.P., and Roberts, R.M. (2012). Model systems for studying trophoblast differentiation from human pluripotent stem cells. Cell and tissue research *349*, 809-824.
Farquharson, R.G., Jauniaux, E., Exalto, N., and Pregnancy, E.S.I.G.f.E. (2005). Updated and revised nomenclature for description of early pregnancy events. Human reproduction *20*, 3008-3011.
Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proceedings of the National Academy of Sciences of the United States of America *105*, 6057-6062.
Founds, S.A., Terhorst, L.A., Conrad, K.P., Hogge, W.A., Jeyabalan, A., and Conley, Y.P. (2011). Gene expression in first trimester preeclampsia placenta. Biological research for nursing *13*, 134-139.
Fournier, T., Therond, P., Handschuh, K., Tsatsaris, V., and Evain-Brion, D. (2008). PPARgamma and early human placental development. Current medicinal chemistry *15*, 3011-3024.
Frum, T., and Ralston, A. (2015). Cell signaling and transcription factors regulating cell fate during formation of the mouse blastocyst. Trends in genetics : TIG *31*, 402-410.

Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J., and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. Genes & development *16*, 784-789.
Genbacev, O., Donne, M., Kapidzic, M., Gormley, M., Lamb, J., Gilmore, J., Larocque, N., Goldfien, G., Zdravkovic, T., McMaster, M.T.*, et al.* (2011). Establishment of human trophoblast progenitor cell lines from the chorion. Stem cells *29*, 1427-1436.
Grabarek, J.B., Zyzynska, K., Saiz, N., Piliszek, A., Frankenberg, S., Nichols, J., Hadjantonakis, A.K., and Plusa, B. (2012). Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. Development *139*, 129-139.
Guillemot, F., Nagy, A., Auerbach, A., Rossant, J., and Joyner, A.L. (1994). Essential role of Mash-2 in extraembryonic development. Nature *371*, 333-336.
Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome biology *8*, R24.
Gurdon, J.B., and Melton, D.A. (2008). Nuclear reprogramming in cells. Science *322*, 1811-1815.
Hailesellasse Sene, K., Porter, C.J., Palidwor, G., Perez-Iratxeta, C., Muro, E.M., Campbell, P.A., Rudnicki, M.A., and Andrade-Navarro, M.A. (2007). Gene function in early mouse embryonic stem cell differentiation. BMC genomics *8*, 85.
Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proceedings of the National Academy of Sciences of the United States of America *107*, 9222-9227.
Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R.*, et al.* (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell *133*, 250-264.
Harvey, K.F., and Hariharan, I.K. (2012). The hippo pathway. Cold Spring Harbor perspectives in biology *4*, a011288.
Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem cells *22*, 225-235.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.
Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nature reviews Molecular cell biology *16*, 144-154.
Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nature reviews Molecular cell biology *10*, 526-537.
Hemberger, M., Udayashankar, R., Tesar, P., Moore, H., and Burton, G.J. (2010). ELF5-enforced transcriptional networks define an epigenetically regulated trophoblast stem cell compartment in the human placenta. Human molecular genetics *19*, 2456-2467.
Herrscher, R.F., Kaplan, M.H., Lelsz, D.L., Das, C., Scheuermann, R., and Tucker, P.W. (1995). The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. Genes & development *9*, 3067-3082.
Hirai, H., Karian, P., and Kikyo, N. (2011). Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. The Biochemical journal *438*, 11-23.

Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. The Journal of biological chemistry *284*, 28729-28737.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.
Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375-386.
Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Molecular cell *62*, 79-91.
Jedrusik, A., Parfitt, D.E., Guo, G., Skamagki, M., Grabarek, J.B., Johnson, M.H., Robson, P., and Zernicka-Goetz, M. (2008). Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. Genes & development *22*, 2692-2706.
Johnson, M.H. (2009). From mouse egg to mouse embryo: polarities, axes, and tissues. Annual review of cell and developmental biology *25*, 483-512.
Johnson, M.H., and Ziomek, C.A. (1981). The foundation of two distinct cell lineages within the mouse morula. Cell *24*, 71-80.
Kelly, S.J. (1977). Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. The Journal of experimental zoology *200*, 365-376.
Kidder, B.L., and Palmer, S. (2010). Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. Genome research *20*, 458-472.
Kidder, B.L., and Palmer, S. (2012). HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. Nucleic acids research *40*, 2925-2939.
Kim, D., and Tucker, P.W. (2006). A regulated nucleocytoplasmic shuttle contributes to Bright's function as a transcriptional activator of immunoglobulin genes. Molecular and cellular biology *26*, 2187-2201.
Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nature protocols *4*, 506-517.
Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.
Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell *143*, 313-324.
Kim, P.G., Canver, M.C., Rhee, C., Ross, S.J., Harriss, J.V., Tu, H.C., Orkin, S.H., Tucker, H.O., and Daley, G.Q. (2016). Interferon-alpha signaling promotes embryonic HSC maturation. Blood.
Kloosterman, W.P., and Plasterk, R.H. (2006). The diverse functions of microRNAs in animal development and disease. Developmental cell *11*, 441-450.
Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics *23*, 1289-1291.
Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. Development *126*, 723-732.

Kubaczka, C., Senner, C.E., Cierlitza, M., Arauzo-Bravo, M.J., Kuckenberg, P., Peitz, M., Hemberger, M., and Schorle, H. (2015). Direct Induction of Trophoblast Stem Cells from Murine Fibroblasts. Cell stem cell *17*, 557-568.
Kuckenberg, P., Buhl, S., Woynecki, T., van Furden, B., Tolkunova, E., Seiffe, F., Moser, M., Tomilin, A., Winterhager, E., and Schorle, H. (2010). The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. Molecular and cellular biology *30*, 3310-3320.
Kuckenberg, P., Peitz, M., Kubaczka, C., Becker, A., Egert, A., Wardelmann, E., Zimmer, A., Brustle, O., and Schorle, H. (2011). Lineage conversion of murine extraembryonic trophoblast stem cells to pluripotent stem cells. Molecular and cellular biology *31*, 1748-1756.
Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes & development *9*, 1250-1262.
Latos, P.A., Sienerth, A.R., Murray, A., Senner, C.E., Muto, M., Ikawa, M., Oxley, D., Burge, S., Cox, B.J., and Hemberger, M. (2015). Elf5-centered transcription factor hub controls trophoblast stem cell self-renewal and differentiation through stoichiometry-sensitive shifts in target gene networks. Genes & development *29*, 2435-2448.
Lessard, J.A., and Crabtree, G.R. (2010). Chromatin regulatory mechanisms in pluripotency. Annual review of cell and developmental biology *26*, 503-532.
Levine, R.J., Lam, C., Qian, C., Yu, K.F., Maynard, S.E., Sachs, B.P., Sibai, B.M., Epstein, F.H., Romero, R., Thadhani, R.*, et al.* (2006). Soluble endoglin and other circulating antiangiogenic factors in preeclampsia. The New England journal of medicine *355*, 992-1005.
Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nature reviews Genetics *3*, 662-673.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
Li, Y., Moretto-Zita, M., Soncin, F., Wakeland, A., Wolfe, L., Leon-Garcia, S., Pandian, R., Pizzo, D., Cui, L., Nazor, K.*, et al.* (2013). BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a DeltaNp63+ cytotrophoblast stem cell state. Development *140*, 3965-3976.
Li, Y., and Parast, M.M. (2014). BMP4 regulation of human trophoblast development. The International journal of developmental biology *58*, 239-246.
Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D.*, et al.* (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nature cell biology *10*, 731-739.
Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H.*, et al.* (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558-562.
Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J.*, et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics *38*, 431-440.
Lu, C.W., Yabuuchi, A., Chen, L., Viswanathan, S., Kim, K., and Daley, G.Q. (2008). Ras-MAPK signaling promotes trophectoderm formation from embryonic stem cells and mouse embryos. Nature genetics *40*, 921-926.
Maekawa, M., Yamamoto, T., Tanoue, T., Yuasa, Y., Chisaka, O., and Nishida, E. (2005). Requirement of the MAP kinase signaling pathways for mouse preimplantation development. Development *132*, 1773-1783.

Martinez-Fierro, M.L., Reyes-Oliva, E.A., Cabral-Pacheco, G.A., Garza-Veloz, I., Aceves-Medina, M.C., Luevano, M., Barbosa-Cisneros, O.Y., Galvan-Valencia, M., Yahuaca-Mendoza, P., Delgado-Enciso, I.*, et al.* (2016). MYC-induced nuclear antigen (MINA) and preeclampsia. Hypertension in pregnancy, 1-13.
McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.
Mikheev, A.M., Nabekura, T., Kaddoumi, A., Bammler, T.K., Govindarajan, R., Hebert, M.F., and Unadkat, J.D. (2008). Profiling gene expression in human placentae of different gestational ages: an OPRU Network and UW SCOR Study. Reproductive sciences *15*, 866-877.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.
Moignard, V., and Gottgens, B. (2014). Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. BioEssays : news and reviews in molecular, cellular and developmental biology *36*, 419-426.
Molineris, I., Grassi, E., Ala, U., Di Cunto, F., and Provero, P. (2011). Evolution of promoter affinity for transcription factors in the human lineage. Molecular biology and evolution *28*, 2173-2183.
Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. Human molecular genetics *14 Spec No 1*, R47-58.
Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S., and Parmacek, M.S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & development *12*, 3579-3590.
Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. Genes & development *12*, 2403-2412.
Niakan, K.K., and Eggan, K. (2013). Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. Developmental biology *375*, 54-64.
Niakan, K.K., Han, J., Pedersen, R.A., Simon, C., and Pera, R.A. (2012). Human pre-implantation embryo development. Development *139*, 829-841.
Niakan, K.K., Ji, H., Maehr, R., Vokes, S.A., Rodolfa, K.T., Sherwood, R.I., Yamaki, M., Dimos, J.T., Chen, A.E., Melton, D.A.*, et al.* (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes & development *24*, 312-326.
Niakan, K.K., Schrode, N., Cho, L.T., and Hadjantonakis, A.K. (2013). Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. Nature protocols *8*, 1028-1041.
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.
Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N.*, et al.* (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Developmental cell *16*, 398-410.

Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mechanisms of development *125*, 270-283.
Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y.*, et al.* (2009). Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell stem cell *5*, 420-433.
Niwa, H. (2007). How is pluripotency determined and maintained? Development *134*, 635-646.
Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes & development *12*, 2048-2060.
Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics *24*, 372-376.
Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.
Okada, Y., Ueshin, Y., Isotani, A., Saito-Fujita, T., Nakashima, H., Kimura, K., Mizoguchi, A., Oh-Hora, M., Mori, Y., Ogata, M.*, et al.* (2007). Complementation of placental defects and embryonic lethality by trophoblast-specific lentiviral gene transfer. Nature biotechnology *25*, 233-237.
Orkin, S.H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T.W., Shen, X., and Levasseur, D.N. (2008). The transcriptional network controlling pluripotency in ES cells. Cold Spring Harbor symposia on quantitative biology *73*, 195-202.
Palmieri, S.L., Peter, W., Hess, H., and Scholer, H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Developmental biology *166*, 259-267.
Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Molecular and cellular biology *27*, 3769-3779.
Pauken, C.M., and Capco, D.G. (1999). Regulation of cell adhesion during embryonic compaction of mammalian embryos: roles for PKC and beta-catenin. Molecular reproduction and development *54*, 135-144.
Perez-Sepulveda, A., Torres, M.J., Khoury, M., and Illanes, S.E. (2014). Innate immune system and preeclampsia. Frontiers in immunology *5*, 244.
Pfeffer, P.L., and Pearton, D.J. (2012). Trophoblast development. Reproduction *143*, 231-246.
Plusa, B., Frankenberg, S., Chalmers, A., Hadjantonakis, A.K., Moore, C.A., Papalopulu, N., Papaioannou, V.E., Glover, D.M., and Zernicka-Goetz, M. (2005). Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. Journal of cell science *118*, 505-515.
Popowski, M., Templeton, T.D., Lee, B.K., Rhee, C., Li, H., Miner, C., Dekker, J.D., Orlanski, S., Bergman, Y., Iyer, V.R.*, et al.* (2014). Bright/Arid3A Acts as a Barrier to Somatic Cell Reprogramming through Direct Regulation of Oct4, Sox2, and Nanog. Stem cell reports *2*, 26-35.
Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development *137*, 395-403.
Ralston, A., and Rossant, J. (2005). Genetic regulation of stem cell origins in the mouse embryo. Clinical genetics *68*, 106-112.

Rayon, T., Menchero, S., Nieto, A., Xenopoulos, P., Crespo, M., Cockburn, K., Canon, S., Sasaki, H., Hadjantonakis, A.K., de la Pompa, J.L.*, et al.* (2014). Notch and hippo converge on Cdx2 to specify the trophectoderm lineage in the mouse blastocyst. Developmental cell *30*, 410-422.
Red-Horse, K., Zhou, Y., Genbacev, O., Prakobphol, A., Foulk, R., McMaster, M., and Fisher, S.J. (2004). Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. The Journal of clinical investigation *114*, 744-754.
Reik, W., Santos, F., Mitsuya, K., Morgan, H., and Dean, W. (2003). Epigenetic asymmetry in the mammalian zygote and early embryo: relationship to lineage commitment? Philosophical transactions of the Royal Society of London Series B, Biological sciences *358*, 1403-1409; discussion 1409.
Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nature biotechnology *18*, 399-404.
Rhee, C., Lee, B.K., Beck, S., Anjum, A., Cook, K.R., Popowski, M., Tucker, H.O., and Kim, J. (2014). Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. Genes & development *28*, 2219-2232.
Riley, P., Anson-Cartwright, L., and Cross, J.C. (1998). The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. Nature genetics *18*, 271-275.
Roberts, J.M., Taylor, R.N., Musci, T.J., Rodgers, G.M., Hubel, C.A., and McLaughlin, M.K. (1989). Preeclampsia: an endothelial cell disorder. American journal of obstetrics and gynecology *161*, 1200-1204.
Roberts, R.M., and Fisher, S.J. (2011). Trophoblast stem cells. Biology of reproduction *84*, 412-421.
Rossant, J. (2004). Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. Seminars in cell & developmental biology *15*, 573-581.
Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527-531.
Rossant, J., and Cross, J.C. (2001). Placental development: lessons from mouse mutants. Nature reviews Genetics *2*, 538-548.
Rossant, J., and Tam, P.P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development *136*, 701-713.
Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Paldi, A., and Viegas-Pequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes & development *12*, 2108-2113.
Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proceedings of the National Academy of Sciences of the United States of America *107*, 10783-10790.
Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C.*, et al.* (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature *404*, 95-99.
Santos, F., and Dean, W. (2004). Epigenetic reprogramming during early development in mammals. Reproduction *127*, 643-651.
Sarkar, P., Randall, S.M., Collier, T.S., Nero, A., Russell, T.A., Muddiman, D.C., and Rao, B.M. (2015). Activin/nodal signaling switches the terminal fate of human embryonic stem cell-derived trophoblasts. The Journal of biological chemistry *290*, 8834-8848.
Scott, I.C., Anson-Cartwright, L., Riley, P., Reda, D., and Cross, J.C. (2000). The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. Molecular and cellular biology *20*, 530-541.

Shimosato, D., Shiki, M., and Niwa, H. (2007). Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. BMC developmental biology *7*, 80.
Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS biology *6*, e253.
Simister, N.E., and Story, C.M. (1997). Human placental Fc receptors and the transmission of antibodies from mother to fetus. Journal of reproductive immunology *37*, 1-23.
Smith, C.M., Finger, J.H., Hayamizu, T.F., McCright, I.J., Xu, J., Berghout, J., Campbell, J., Corbani, L.E., Forthofer, K.L., Frost, P.J.*, et al.* (2014). The mouse Gene Expression Database (GXD): 2014 update. Nucleic acids research *42*, D818-824.
Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555-568.
Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell stem cell *2*, 230-240.
Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development *132*, 2093-2102.
Suwinska, A., Czolowska, R., Ozdzenski, W., and Tarkowski, A.K. (2008). Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. Developmental biology *322*, 133-144.
Suzuki, S. (2008). Clinical significance of pregnancies with circumvallate placenta. The journal of obstetrics and gynaecology research *34*, 51-54.
Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.
Takaoka, K., and Hamada, H. (2012). Cell fate decisions and axis determination in the early mouse embryo. Development *139*, 3-14.
Tam, P.P., and Rossant, J. (2003). Mouse embryonic chimeras: tools for studying mammalian development. Development *130*, 6155-6163.
Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. Science *282*, 2072-2075.
Tarkowski, A.K. (1959). Experiments on the development of isolated blastomers of mouse eggs. Nature *184*, 1286-1287.
Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.
Urvalek, A.M., and Gudas, L.J. (2014). Retinoic acid and histone deacetylases regulate epigenetic changes in embryonic stem cells. The Journal of biological chemistry *289*, 19519-19530.
Uy, G.D., Downs, K.M., and Gardner, R.L. (2002). Inhibition of trophoblast stem cell potential in chorionic ectoderm coincides with occlusion of the ectoplacental cavity in the mouse. Development *129*, 3913-3924.
Waite, L.L., Person, E.C., Zhou, Y., Lim, K.H., Scanlan, T.S., and Taylor, R.N. (2000). Placental peroxisome proliferator-activated receptor-gamma is up-regulated by pregnancy serum. The Journal of clinical endocrinology and metabolism *85*, 3808-3814.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.
Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F.*, et al.* (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621-635.
Webb, C.F., Bryant, J., Popowski, M., Allred, L., Kim, D., Harriss, J., Schmidt, C., Miner, C.A., Rose, K., Cheng, H.L.*, et al.* (2011). The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. Molecular and cellular biology *31*, 1041-1053.
Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J., Schafer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H.*, et al.* (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biology of reproduction *82*, 214-223.
Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A.*, et al.* (2003). Database resources of the National Center for Biotechnology. Nucleic acids research *31*, 28-33.
Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.
Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.
Wilmut, I., Beaujean, N., de Sousa, P.A., Dinnyes, A., King, T.J., Paterson, L.A., Wells, D.N., and Young, L.E. (2002). Somatic cell nuclear transfer. Nature *419*, 583-586.
Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd*, et al.* (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome biology *10*, R130.
Wu, G., Gentile, L., Fuchikami, T., Sutter, J., Psathaki, K., Esteves, T.C., Arauzo-Bravo, M.J., Ortmeier, C., Verberk, G., Abe, K.*, et al.* (2010). Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. Development *137*, 4159-4169.
Wu, T., Wang, H., He, J., Kang, L., Jiang, Y., Liu, J., Zhang, Y., Kou, Z., Liu, L., Zhang, X.*, et al.* (2011). Reprogramming of trophoblast stem cells into pluripotent stem cells by Oct4. Stem cells *29*, 755-763.
Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nature biotechnology *20*, 1261-1264.
Yagi, R., Kohn, M.J., Karavanova, I., Kaneko, K.J., Vullhorst, D., DePamphilis, M.L., and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development *134*, 3827-3836.
Yamauchi, N., Kiessling, A.A., and Cooper, G.M. (1994). The Ras/Raf signaling pathway is required for progression of mouse embryos through the two-cell stage. Molecular and cellular biology *14*, 6655-6662.
Yang, Y., Bai, W., Zhang, L., Yin, G., Wang, X., Wang, J., Zhao, H., Han, Y., and Yao, Y.Q. (2008). Determination of microRNAs in mouse preimplantation embryos by microarray. Developmental dynamics : an official publication of the American Association of Anatomists *237*, 2315-2327.

Yeap, L.S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. Epigenetics & chromatin *2*, 12.
Yu, J., and Thomson, J.A. (2008). Pluripotent stem cell lines. Genes & development *22*, 1987-1997.
Yu, P., Pan, G., Yu, J., and Thomson, J.A. (2011). FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. Cell stem cell *8*, 326-334.
Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.H., Yaw, L.P., Robson, P., Lim, B., and Ng, H.H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes & development *23*, 2507-2520.
Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & development *25*, 2227-2241.
Zeisler, H., Hund, M., and Verlohren, S. (2016). The sFlt-1:PlGF Ratio in Women with Suspected Preeclampsia. The New England journal of medicine *374*, 1785-1786.
Zernicka-Goetz, M. (2004). First cell fate decisions and spatial patterning in the early mouse embryo. Seminars in cell & developmental biology *15*, 563-572.
Zernicka-Goetz, M., Morris, S.A., and Bruce, A.W. (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. Nature reviews Genetics *10*, 467-477.
Zhang, J.Z., Gao, W., Yang, H.B., Zhang, B., Zhu, Z.Y., and Xue, Y.F. (2006). Screening for genes essential for mouse embryonic stem cell self-renewal using a subtractive RNA interference library. Stem cells *24*, 2661-2668.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W*., et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.
Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X*., et al.* (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Molecular cell *63*, 1066-1079.

VITA

Catherine Rhee was born in Hartford, Connecticut. She entered University of Wisconsin-Madison in 2007 and received Bachelor of Science in Microbiology and Genetics in 2011. In June of 2011, she started her graduate study under supervision of Dr. Jonghwan Kim and Dr. Haley Tucker in the Cell and Molecular Biology program at the University of Texas at Austin.

Email: catherinerhee@mail.utexas.edu

This dissertation was typed by the author, Catherine Soo Rhee.